Nonparametric lack-of-fit tests for parametric mean-regression models with censored data

Olivier Lopez, Valentin Patilea

Nonparametric lack-of-fit tests for parametric mean-regression models with censored data

O. Lopez*   V. Patilea†

March 30, 2008

Abstract

We develop two kernel smoothing based tests of a parametric mean-regression model against a nonparametric alternative when the response variable is right-censored. The new test statistics are inspired by the synthetic data and the weighted least squares approaches for estimating the parameters of a (non)linear regression model under censoring. The asymptotic critical values of our tests are given by the quantiles of the standard normal law. The tests are consistent against fixed alternatives, local Pitman alternatives and uniformly over alternatives in Hölder classes of functions of known regularity.

Key words: Hypothesis testing, censored data, Kaplan-Meier integral, local alternative

MSC 2000: 62G10, 62G08, 62N01

*IRMAR and CREST-ENSAI. Email: lopez@ensai.fr. Corresponding author. Address for correspondence: CREST-ENSAI, Campus de Ker Lann, Rue Blaise Pascal BP 37203, 35172 Bruz cedex, FRANCE

†CREST-ENSAI and Insa-IRMAR. Email: patilea@ensai.fr

1 Introduction

Parametric mean-regression models, in particular the linear model, are valuable tools for exploring the relationship between a response and a set of explanatory variables (covariates). However, in survival analysis such models are overshadowed by the fashionable proportional hazard models and the accelerated failure time models where one imposes a form for the conditional law of the response given the covariates. Even though mean-regression models involve weaker assumptions on the conditional law of the responses, the popularity of the parametric mean-regressions with censored data greatly suffers from the difficulty to perform statistical inference when not all responses are available.

The existing methods for the estimation of the parameters of the mean-regression in the presence of right censoring can be split into two main categories: i) *weighted least squares (WLS)* based on the uncensored observations but suitably weighted to account for censorship (see Zhou 1992, Stute 1999); and ii) *synthetic data (SD)* estimators obtained by ordinary least squares with transformed responses, using a transformation that preserves the conditional expectation and that can be estimated from data (e.g., Koul *et al.* 1981, Leurgans 1987).

This paper's main purpose focuses on a further step in the statistical inference for parametric mean-regression models under right censoring, that is nonparametric lack-of-fit testing. Checking the adequacy of a parametric regression function against a purely nonparametric alternative has received a large amount of attention in the non-censored case and several approaches have been proposed. See, amongst many others, Härdle and Mammen (1993), Zheng (1996), Stute (1997), Horowitz and Spokoiny (2001), Guerre and Lavergne (2005), and the references therein. But for right-censored data, these approaches are not directly applicable. To our knowledge, very few solutions for nonparametric regression checks with right-censored responses have been proposed. Following the approach of Stute (1997), Stute *et al.* (2000) introduced two tests based on an empirical process marked by *weighted* residuals, the role of the weights being to account for censoring. The limit of their marked empirical process is a rather complicated centered Gaussian process and therefore the implementation of the test requires numerical calculations. Sánchez-Sellero *et al.* (2005) reconsidered this type of test and provided a complete proof of its asymptotic level. However, for technical reasons, Sánchez-Sellero *et al.* (2005) drop some observations in the right tail of the response variable and therefore the resulting tests *are no longer omnibus*. Moreover, neither Stute *et al.* (2000) nor Sánchez-Sellero *et al.* (2005) studied the consistency of the tests against a sequence of alternatives approaching the null hypothesis. Pardo-Fernandez *et al.* (2005) proposed another test for parametric models in censored regression that is based on the comparison of two estimators, parametric and nonparametric, of the distribution of the errors. As the latter estimator is based on a nonparametric location-scale model, the test of Pardo-Fernandez *et al.* (2005) is not consistent against *any* alternative.

In this paper we consider two versions adapted for right-censored responses of the kernel-based test statistic studied by Zheng (1996). See also Härdle and Mammen (1993), Horowitz and Spokoiny (2001), Guerre and Lavergne (2005) for closely related test statistics. In the non-censored case, the kernel-based test statistic we consider is a suitably normalized $U-$statistic built from the estimated residuals of the parametric model. Under suitable conditions, the test statistic converges in law to a standard normal when the model is correct. The problem in presence of censoring is that estimated residuals can be computed only for uncensored observations. The two solutions we propose are inspired by the WLS and SD estimation approaches mentioned above. On one hand, we build a *weighted* $U-$statistic using estimated residuals with the weights estimated from data. Once again, the weights account for censoring. On the other hand, we build a $U-$statistic using estimated *synthetic* residuals where the synthetic residuals are the difference between the synthetic responses and the predictions given by the model. Two smoothing-based test statistics are obtained after suitably normalizing each of these $U-$statistics.

The paper is organized as follows. In section 2 we recall the weighted least squares and synthetic data approaches for (non)linear regression models when the response is right-censored. Section 3 shows how to build two kernel based test statistics adapted for censored responses. Section 4 deals with the asymptotic behavior of the two omnibus tests that we derive. The main results in this paper show that the asymptotic study of our tests boils down to the asymptotic study of kernel-based tests without censoring but with suitably transformed observations. As a consequence, the asymptotic critical values of the new tests are given by the quantiles of the standard normal law. Moreover, the asymptotic consistency of our tests is obtained by arguments similar to those used for kernel based tests in the non-censored case. In particular, we study the consistency of the new tests against fixed alternatives, local Pitman type alternatives and the consistency uniformly over Hölder classes of alternatives of known regularity. The performances of the kernel-based tests we propose depend on the choice of the bandwidth. Inspired by the maximum test approach of Horowitz and Spokoiny (2001), we propose a data-driven procedure to select the bandwidth with censored responses. However, to keep this paper at reasonable length, the detailed theoretical and empirical investigation of this data-driven procedure is left for future work. Finally, in section 5 we illustrate the performance of the new tests using simulated and real data.

2 Preliminaries

Consider the model $Y = m(X) + \varepsilon$, where $Y \in \mathbb{R}$, $X \in \mathbb{R}^p$, $\mathbb{E}(\varepsilon \mid X) = 0$ almost surely (a.s.), and $m(\cdot)$ is an unknown function. In presence of random right censoring, the response Y is not always available. Instead of (Y, X), one observes a random sample from (T, δ, X) with

$$T = Y \wedge C, \qquad \delta = \mathbf{1}_{\{Y \leq C\}},$$

where C is the "censoring" random variable, and $\mathbf{1}_A$ denotes the indicator function of the set A. In our setting, the variable X is not subject to censoring and is fully observed. We want to check whether the regression function $m(\cdot)$ belongs to a parametric family $\mathcal{M} = \left\{f(\theta, \cdot) : \theta \in \Theta \subset \mathbb{R}^d\right\}$ where f is a known function. Our null hypothesis then writes

$$H_0 \ : \text{for some } \theta_0, \quad \mathbb{E}(Y|X) = f(\theta_0, X) \text{ a.s.}, \tag{2.1}$$

while the alternative is $\mathbb{P}[\mathbb{E}(Y|X) = f(\theta, X)] \leq c$ for every $\theta \in \Theta$ and some $c < 1$. For testing H_0, first we need to estimate θ_0.

2.1 Estimating (non)linear regressions with censored data

Since the observed variable T does not have the same conditional expectation as Y, classical techniques for estimating parametric (non)linear regression models like $\mathcal{M}$ must be adapted to account for censorship. Several adapted procedures have been proposed, that we classify in two groups: synthetic data (SD) procedures and weighted least squares (WLS). In the SD approach one replaces the variable T with some transformation of the data Y^*, a transformation which preserves the conditional expectation of Y. Several transformations have been proposed, see for instance Leurgans (1987), Zheng (1987). In the following, we will restrain ourselves to the transformation first proposed by Koul *et al.* (1981), that is

$$Y^* = \frac{\delta T}{1 - G\,(T-)}, \tag{2.2}$$

where $G\,(t) = \mathbb{P}\,(C \leq t)$. The following assumptions will be used throughout this paper to ensure that $\mathbb{E}\,(Y^* \mid X) = \mathbb{E}\,(Y \mid X)$ for Y^* defined in (2.2).

Assumption 1 *Y and C are independent.*

Assumption 2 $\mathbb{P}\,(Y \leq C \mid X, Y) = \mathbb{P}\,(Y \leq C \mid Y)\,.$

These assumptions are quite common in the survival analysis literature when covariates are present. Assumption 1 is an usual identification condition when working with the Kaplan-Meier estimator. Stute (1993), pages 462-3, provides a detailed discussion on Assumption 2. These assumptions may be inappropriate for some data sets. However, they are often satisfied in randomized clinical trials when the failure time Y of each subject is either observed or administratively censored at the end of the follow-up period. Notice that Assumption 2 is flexible enough to allow for a dependence between X and C. Moreover, Assumptions 1 and 2 imply the following general property: for any integrable $\phi(T, X)$,

$$\mathbb{E}\left[\frac{\delta}{1 - G(T-)}\,\phi(T, X) \mid X\right] = \mathbb{E}\left[\phi(Y, X) \mid X\right]. \tag{2.3}$$

Unfortunately, one cannot compute the transformation (2.2) when the function G is unknown. Given the i.i.d. sample $(T_1, \delta_1, X_1)\,, ..., (T_n, \delta_n, X_n)\,,$ Koul *et al.* (1981) proposed to replace G with its Kaplan-Meier estimate

$$\hat{G}\,(t) = 1 - \prod_{\{j : T_j \leq t\}} \left(1 - \frac{1}{R_n\,(T_j)}\right)^{1-\delta_j}, \qquad \text{with} \qquad R_n\,(t) = \sum_{k=1}^{n} \mathbf{1}_{\{t \leq T_k\}},$$

and to compute

$$\hat{Y}_i^* = \frac{\delta_i T_i}{1 - \hat{G}\,(T_i-)}, \quad i = 1, ..., n. \tag{2.4}$$

Next, Koul *et al.* (1981) proposed to estimate θ_0 by $\hat{\theta}^{SD}$ that minimizes

$$M_n^{SD}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{Y}_i^* - f(\theta, X_i) \right]^2$$

over Θ. They obtained the consistency of $\hat{\theta}^{SD}$ and the asymptotic normality of $\sqrt{n}(\hat{\theta}^{SD} - \theta_0)$ in the particular case of a linear regression model. Delecroix *et al.* (2006) generalized these results to more general functions $f(\theta, x)$.

The WLS approach consists of applying weighted least squares techniques directly to variables T_i, that is computing $\hat{\theta}^{WLS}$ which minimizes

$$M_n^{WLS}(\theta) = \sum_{i=1}^{n} W_{in} \left[T_i - f(\theta, X_i) \right]^2,$$

with a specific choice of W_{in} that compensates for the fact that Y is censored. More precisely, the weights W_{in} are defined by

$$W_{in} = \frac{\delta_i}{n \left[1 - \hat{G}(T_i-) \right]}. \tag{2.5}$$

Zhou (1992) studied an estimator like $\hat{\theta}^{WLS}$ in the case of linear regression. Under Assumptions 1 and 2, Stute (1999) generalized this approach to nonlinear regressions. Using the Kaplan-Meier estimator $\hat{F}_{(X,Y)}(x, y)$ of $F_{(X,Y)}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ introduced by Stute (1993), Stute (1999) interpreted $\hat{\theta}^{WLS}$ as the minimizer of

$$\int \left[y - f(\theta, x) \right]^2 d\hat{F}_{(X,Y)}(x, y) \tag{2.6}$$

with respect to θ. Indeed, on one hand, by definition, at observation i the jump of $\hat{F}_{(X,Y)}$ is equal to the jump of the Kaplan-Meier estimate of $F(t) = \mathbb{P}(Y \leq t)$. On the other hand, it can be easily shown that the jump of $\hat{F}(t)$ at observation i is equal to the weight W_{in} defined in (2.5). Using the properties of Kaplan-Meier integrals, one can deduce consistency and $\sqrt{n}-$asymptotic normality for $\hat{\theta}^{WLS}$. See Stute (1999, 1993) or Delecroix *et al.* (2006). It is worthwhile to notice that a choice of W_{in} as in (2.5) connects $M_n^{WLS}(\theta)$ to $M_n^{SD}(\theta)$ since $\hat{Y}_i^* = nW_{in}T_i$. In the following section, we extend the purpose of the SD and WLS methodologies from estimation to testing.

3 Nonparametric test procedures under censoring

To better explain the new approach, first the case where Y is not censored is reconsidered. Then, testing the adequacy of model $\mathcal{M}$ is equivalent to testing

$$\text{for some } \theta_0, \quad Q(\theta_0) = 0 \quad \text{where} \quad Q(\theta) = \mathbb{E}\left[U(\theta)\,\mathbb{E}\left[U(\theta) \mid X\right] g(X)\right],$$

$U(\theta) = Y - f(\theta, X)$ and g denotes the density of X that is assumed to exist. The choice of g avoids handling denominators close to zero. When the responses are *not censored*, one may estimate $Q(\theta_0)$ by

$$Q_n(\hat{\theta}) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} U_i(\hat{\theta}) U_j(\hat{\theta}) K_h(X_i - X_j) \tag{3.1}$$

where $\hat{\theta}$ is an estimator of θ_0 such that $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$, $U_i(\theta) = Y_i - f(\theta, X_i)$, K is some $p-$dimensional kernel function, h denotes the bandwidth and for $x \in \mathbb{R}^p$, $K_h(x) = K(x/h)$. See Zheng (1996). See also Horowitz and Spokoiny (2001) or Guerre and Lavergne (2005).

Using a consistent estimate $\hat{V}_n^2$ of the asymptotic variance of $nh^{p/2}Q_n(\hat{\theta})$, the smoothing based test statistic with non-censored responses is

$$T_n^{NC} = nh^{p/2}\frac{Q_n(\hat{\theta})}{\hat{V}_n}. \tag{3.2}$$

Under the null hypothesis the statistic behaves asymptotically as a standard normal and therefore the nonparametric test is defined as *"Reject H_0 when $T_n^{NC} \geq z_{1-\alpha}$"*, where $z_{1-\alpha}$ is the $(1-\alpha)$th quantile of the standard normal law. As an estimate $\hat{V}_n^2$, one could use either

$$\hat{V}_n^2 = \frac{2}{n(n-1)h^p} \sum_{i \neq j} U_i^2(\hat{\theta}) U_j^2(\hat{\theta}) K_h^2(X_i - X_j)$$

or

$$\hat{V}_n^2 = \frac{2}{n(n-1)h^p} \sum_{i \neq j} \hat{\sigma}^2(X_i)\,\hat{\sigma}^2(X_j)\,K_h^2(X_i - X_j), \tag{3.3}$$

with $\hat{\sigma}^2(x)$ a nonparametric estimator of $\sigma^2(x) = Var(\varepsilon \mid X = x)$. The former choice for $\hat{V}_n^2$ is simpler but is likely to decrease the power of the test because the squares of the estimated residuals of the parametric model produce an upward biased estimate of $\sigma^2(x)$ under the alternative hypothesis. In the presence of censored responses, the test statistic (3.2) cannot be computed since $U_i(\theta)$ are not available for censored observations.

3.1 Two test statistics with right-censored responses

In the following, the observations are (T_i, δ_i, X_i), $1 \leq i \leq n$, a random sample from (T, δ, X). In the spirit of the SD approach, consider

$$Q_n^{SD}(\hat{\theta}) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} \hat{U}_i^{SD}(\hat{\theta}) \hat{U}_j^{SD}(\hat{\theta}) K_h(X_i - X_j), \tag{3.4}$$

where $\hat{\theta} = \hat{\theta}^{SD}$ and

$$\hat{U}_i^{SD}(\theta) = \frac{\delta_i}{1 - \hat{G}(T_i-)} T_i - f(\theta, X_i) = nW_{in} T_i - f(\theta, X_i) \tag{3.5}$$

are the estimated synthetic residuals. The statistic $Q_n^{SD}(\theta)$ estimates

$$Q^{SD}(\theta) = \mathbb{E}\left[U^{SD}(\theta)\, \mathbb{E}\left[U^{SD}(\theta) \mid X\right] g(X)\right]$$

with $U^{SD}(\theta) = \delta T\left[1 - G(T-)\right]^{-1} - f(\theta, X)$. By (2.3), if Assumptions 1 and 2 hold then the null hypothesis is equivalent to $Q^{SD}(\theta_0) = 0$.

On the other hand, following the WLS approach we can replace $Q_n(\hat{\theta})$ in (3.1) with

$$Q_n^{WLS}(\hat{\theta}) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} \hat{U}_i^{WLS}(\hat{\theta}) \hat{U}_j^{WLS}(\hat{\theta}) K_h(X_i - X_j), \tag{3.6}$$

where $\hat{\theta} = \hat{\theta}^{WLS}$ and

$$\hat{U}_i^{WLS}(\theta) = \frac{\delta_i}{1 - \hat{G}(T_i-)} \left[T_i - f(\theta, X_i)\right] = nW_{in}\left[T_i - f(\theta, X_i)\right]. \tag{3.7}$$

The statistic $Q_n^{WLS}(\theta)$ estimates

$$Q^{WLS}(\theta) = \mathbb{E}\left[U^{WLS}(\theta)\, \mathbb{E}\left[U^{WLS}(\theta) \mid X\right] g(X)\right]$$

with $U^{WLS}(\theta) = \delta\left[1 - G(T-)\right]^{-1}\left[T - f(\theta, X)\right]$. By (2.3), the null hypothesis is equivalent to $Q^{WLS}(\theta_0) = 0$.

Now, given consistent estimates $\left[\hat{V}_n^{SD}\right]^2$ and $\left[\hat{V}_n^{WLS}\right]^2$ of the asymptotic variance of $nh^{p/2}Q_n^{SD}(\hat{\theta})$ and $nh^{p/2}Q_n^{WLS}(\hat{\theta})$, respectively, we introduce

$$T_n^{SD} = T_n^{SD}(\hat{\theta}) = nh^{p/2} \frac{Q_n^{SD}(\hat{\theta})}{\hat{V}_n^{SD}}, \qquad T_n^{WLS} = T_n^{WLS}(\hat{\theta}) = nh^{p/2} \frac{Q_n^{WLS}(\hat{\theta})}{\hat{V}_n^{WLS}}.$$

The corresponding omnibus tests are

$$\text{“Reject } H_0 \text{ when } T_n^{SD} \geq z_{1-\alpha} \quad (\text{resp. } T_n^{WLS} \geq z_{1-\alpha})\text{ ”}. \tag{3.8}$$

To estimate the variance of $nh^{p/2}Q_n^{SD}(\hat{\theta})$ we consider

$$\left[\hat{V}_n^{SD}\right]^2 = \frac{2}{n(n-1)h^p}\sum_{i\neq j}\left[\hat{U}_i^{SD}(\hat{\theta})\right]^2\left[\hat{U}_j^{SD}(\hat{\theta})\right]^2 K_h^2\left(X_i - X_j\right). \tag{3.9}$$

The variance of $nh^{p/2}Q_n^{WLS}(\hat{\theta})$ is estimated similarly with $\hat{U}_i^{SD}(\hat{\theta})$ replaced by $\hat{U}_i^{WLS}(\hat{\theta})$. Alternative variance estimates are discussed in section 4.

Checking the validity of a parametric conditional model has attracted much attention in survival analysis. Hjort (1990) and Lin and Spiekerman (1996) considered goodness-of-fit statistics based on martingale residuals, while Gray and Pierce (1985) showed how Neyman's smooth tests may be adapted to censored data. See chapter 10 of Lawless (2003) for a review of the methods for testing the lack-of-fit. All these techniques can be used to check whether some parametric form of the *conditional law* of the response given the explanatory variables is consistent with observed data. Therefore, these techniques are only of limited use in our framework where we aim to check the adequacy of some parametric form of the *conditional expectation* of the response variable given the covariates. The standard normal limit of the test statistics T_n^{SD} and T_n^{WLS} under the null hypothesis, a property that will be proved in the following, yields the simple one-sided tests (3.8) for checking mean-regressions. By contrast, the alternative test statistics available in the literature (see Stute *et al.* 2000) have a complicated limit and there is no simple way to construct the critical values of the associated tests.

4 Asymptotic analysis

The most difficult part of the study of our tests is the investigation of $Q_n^{SD}(\theta)$ and $Q_n^{WLS}(\theta)$. These quadratic forms are difficult to analyze even under H_0 and for $\theta = \theta_0$, since they do not rely on i.i.d. quantities U_i, as the quadratic form (3.1) does. Due to the presence of $\hat{G}$ in (3.5) and (3.7), each $\hat{U}_i^{SD}(\theta_0)$ and $\hat{U}_i^{WLS}(\theta_0)$ depend on the whole sample. Then, a key point is to show that under H_0, in some sense, $Q_n^{SD}(\hat{\theta})$ and $Q_n^{WLS}(\hat{\theta})$ are asymptotically equivalent to the "ideal" quadratic forms

$$\tilde{Q}_n^{SD}(\theta_0) = \frac{1}{n(n-1)h^p}\sum_{i\neq j} U_i^{SD}(\theta_0)\, U_j^{SD}(\theta_0)\, K_h\left(X_i - X_j\right) \tag{4.1}$$

and

$$\tilde{Q}_n^{WLS}(\theta_0) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} U_i^{WLS}(\theta_0) U_j^{WLS}(\theta_0) K_h(X_i - X_j), \tag{4.2}$$

respectively, where

$$\begin{aligned} U_i^{SD}(\theta) &= \frac{\delta_i}{1 - G(T_i-)} T_i - f(\theta, X_i) = \gamma(T_i) T_i - f(\theta, X_i), \\ U_i^{WLS}(\theta) &= \frac{\delta_i}{1 - G(T_i-)} [T_i - f(\theta, X_i)] = \gamma(T_i) [T_i - f(\theta, X_i)]. \end{aligned}$$

The asymptotic study of $\tilde{Q}_n^{SD}(\theta_0)$ and $\tilde{Q}_n^{WLS}(\theta_0)$ can be done like in the i.i.d. non-censored case. See, for instance, Zheng (1996), Horowitz and Spokoiny (2001), Guerre and Lavergne (2005). A similar equivalence result deduced under fixed or moving alternatives will serve for studying the asymptotic consistency of our tests.

4.1 Assumptions

In the following, $\tau_L = \inf\{t \mid L(t) = 1\}$ for any distribution function L.

Assumption 3 *(i) F and G are continuous.*

(ii) $-\infty < \tau_F \leq \tau_G \leq \infty$.

Assumption 3-(i) is introduced for convenience purposes. It allows us to use a simpler i.i.d. representation of the Kaplan-Meier estimator (see Theorem 1 of Major and Rejtő, 1988). Moreover, Assumption 3-(i), considered together with Assumption 1, implies $\mathbb{P}(Y = C) = 0$ and this latter condition justifies the definition of the Kaplan-Meier estimate $\hat{G}$. When $\tau_F > \tau_G$, in general, there is no way to consistently estimate θ_0. Assumption 3-(ii) allows one to avoid this case.

Assumption 4 *(Data): (i) Let $(\varepsilon_1, C_1, X_1), ..., (\varepsilon_n, C_n, X_n)$ be an independent sample of (ε, C, X) where $\varepsilon, C \in \mathbb{R}$ and $X \in \mathbb{R}^p$, and suppose $\mathbb{E}(\varepsilon \mid X) = 0$ a.s.*

(ii) X is a random vector with bounded support $\mathcal{X}$ and bounded density g.

(iii) There exist some constants c_{inf}, c_{sup} such that for each $x \in \mathcal{X}$

$$0 < c_{inf} \leq \mathbb{E}\left[\varepsilon^2 \mid X = x\right] \leq \mathbb{E}\left[\left\{1 + \varepsilon^2\right\}\{1 - G(Y)\}^{-1} \mid X = x\right] \leq c_{sup} < \infty.$$

(iv) $\mathbb{E}\left[\{1 + \varepsilon^4\}\,\delta\{1 - G(Y)\}^{-4}\right] = \mathbb{E}\left[\{1 + \varepsilon^4\}\,\gamma(T)^4\right] < \infty$.

Assumptions 4 (iii)-(iv) are counterparts of assumptions on the conditional variance and the fourth moment of the residuals that are usually imposed in the non-censored case. See, e.g., Guerre and Lavergne (2005). Now, define $\nabla_\theta f(\theta, x) = \partial f(\theta, x)/\partial\theta$, $\nabla_\theta^2 f(\theta, x) = \partial^2 f(\theta, x)/\partial\theta\partial\theta'$, whenever these derivatives exist. For any matrix A, let $\|A\|_2 = \sup_{v\neq 0} \|Av\|/\|v\|$ where $\|v\|$ is the Euclidean norm of the vector v.

Assumption 5 *(Parametric model): The parameter set* Θ *is a compact subset of* $\mathbb{R}^d$, $d \geq 1$, *and* θ_0 *in an interior point of* Θ. *The parametric regression model* $\mathcal{M} = \{f(\theta, \cdot) : \theta \in \Theta\}$ *satisfies:*

(i) Differentiability in θ*: for each* $x \in \mathcal{X}$, $f(\theta, x)$ *is twice differentiable with respect to* θ*. There exists a finite constant* c_1 *such that for each* $\theta \in \Theta$ *and* $x \in \mathcal{X}$, $|f(\theta, x)| + \|\nabla_\theta f(\theta, x)\| + \|\nabla_\theta^2 f(\theta, x)\|_2 \leq c_1$*. Moreover, there exist finite constants* $a, c_2 > 0$ *such that for each* θ *and* x, $|\nabla_\theta^2 f(\theta, x)_{jk} - \nabla_\theta^2 f(\theta_0, x)_{jk}| \leq c_2\|\theta - \theta_0\|^a$, *where* $\nabla_\theta^2 f(\theta, x)_{jk}$ *is the element* jk *of the matrix* $\nabla_\theta^2 f(\theta, x)$.

(ii) Identifiability: there exists a bounded function $\Phi \geq 0$ *with* $\mathbb{E}[\Phi(X)] > 0$ *such that for each* $\theta \in \Theta$ *and* $x \in \mathcal{X}$, $|f(\theta, x) - f(\theta_0, x)| \geq \Phi(x)\|\theta - \theta_0\|$.

Assumption 6 *(Kernel smoother): (i) If* $x = (x_1, ..., x_p)$, *let* $K(x) = \tilde{K}(x_1) ... \tilde{K}(x_p)$ *where* $\tilde{K}$ *is a symmetric continuous density of bounded variation on* $\mathbb{R}$*. The Fourier Transform* $\hat{\tilde{K}}$ *of* $\tilde{K}$ *is positive, integrable and non-increasing on* $[0, \infty)$.

(ii) The bandwidth h *belongs to an interval* $\mathcal{H}_n = [h_{min}, h_{max}]$, $n \geq 1$, *such that* $h_{max} \to 0$ *and* $nh_{min}^{3p} \to \infty$.

Assumption 6-(i) holds, for instance, for normal, Laplace or Cauchy densities. The condition non-increasing Fourier Transform for $\hat{\tilde{K}}$ will serve only for deriving our asymptotic equivalence results *uniformly* in the bandwidth (see, for instance, the proof of Lemma A.7 in the Appendix). Concerning the range for the bandwidth, in view of equation (A.6) in the Appendix, it is clear that h_{min} may be taken of smaller rate if Assumption 4 (iv) above and Assumption 7 below are made more restrictive. The following assumption will allow to control the jumps of the Kaplan-Meier estimator; see also condition (1.6) of Stute (1995) and Stute (1996). Below, $a \vee b$ denotes the maximum of a and b.

Assumption 7 *Let* $q_\rho(x) = \mathbb{E}\left[\{|Y|+1\}C(Y)^{1/2+\rho} \mid X = x\right]$ *where*

$$C(y) = \int_{-\infty}^{y} \frac{dG(t)}{[1-H(t)][1-G(t)]} \vee 1, \quad y \in \mathbb{R},$$

with $H(t) = \mathbb{P}(T \leq t)$. *Then* $\mathbb{E}[q_\rho^2(X)] < \infty$ *for some* $0 < \rho < 1/2$.

The function $C(\cdot)$ also appears in Bose and Sen (2002) who derive an i.i.d. representation for Kaplan-Meier $U-$statistics that would have been useful for deriving our test results. Unfortunately, they impose $\rho = 1/2$ (see Bose and Sen's Theorem 1 and Remark 1) which is unrealistic in our framework.

4.2 Behavior of the tests under the null hypothesis

The following theorem gives an asymptotic representation of the statistics T_n^{SD} and T_n^{WLS} under H_0 stated in (2.1). The proof is postponed to the Appendix. To simplify notation, below we replace the superscripts SD and WLS with 0 and 1, respectively. For instance, we write Q_n^0 (resp. Q_n^1) instead of Q_n^{SD} (resp. Q_n^{WLS}). As before, $\hat\theta$ denotes $\hat\theta^{SD}$ or $\hat\theta^{WLS}$.

Theorem 4.1 *Let Assumptions 1 to 7 hold. Under* H_0, *for* $\beta = 0$ *or 1*

$$\sup_{h\in\mathcal{H}_n}\left\{\left|nh^{p/2}Q_n^\beta(\hat\theta) - nh^{p/2}\tilde{Q}_n^\beta(\theta_0)\right| + \left|\frac{\tilde{V}_n^\beta(\theta_0)}{\hat{V}_n^\beta} - 1\right|\right\} \to 0,$$

in probability, where

$$\left[\tilde{V}_n^\beta(\theta_0)\right]^2 = \frac{2}{n(n-1)h^p}\sum_{i\neq j}\left[U_i^\beta(\theta_0)\right]^2\left[U_j^\beta(\theta_0)\right]^2 K_h^2\left(X_i - X_j\right).$$

Moreover, under H_0 *and for* $\beta = 0$ *or 1*

$$\sup_{h\in\mathcal{H}_n}\left|T_n^\beta(\hat\theta) - \frac{nh^{p/2}\tilde{Q}_n^\beta(\theta_0)}{\tilde{V}_n^\beta(\theta_0)}\right| = o_P(1).$$

Corollary 4.2 *Under Assumptions 1 to 7 the two tests defined in equation (3.8) have asymptotic level* α.

Remark 1. To estimate the variance $nh^{p/2}Q_n^0(\hat\theta)$ we considered (3.9). Alternatively, extending the idea behind (3.3) to the censoring framework, one may replace in (3.9) the estimated squared residual $\hat{U}_i^0(\hat\theta)^2$ with a nonparametric estimate of $\sigma^{*2}(x) = Var(Y^* \mid$

$X = x$). It is easy to check that $Var(Y^* \mid X) = \mathbb{E}\left[U^0\left(\theta_0\right)^2 \mid X\right]$ under H_0 and, in general, $Var(Y^* \mid X) < \mathbb{E}\left[U^0\left(\theta_0\right)^2 \mid X\right]$ if the model $\mathcal{M}$ is wrong. To estimate $\sigma^{*2}(\cdot)$, one can use

$$\hat{\sigma}_n^{*2}(x) = \frac{\sum_{i=1}^n \hat{Y}_i^{*2} L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} - \left(\frac{\sum_{i=1}^n \hat{Y}_i^{*} L((X_i - x)/b_n)}{\sum_{i=1}^n L((X_i - x)/b_n)} \right)^2, \tag{4.3}$$

$x \in \mathcal{X}$, with L a kernel and b_n a bandwidth chosen independently of $\mathcal{H}_n$. If

$$\sup_{x \in \mathcal{X}} \left| \hat{\sigma}_n^{*2}(x) - \sigma^{*2}(x) \right| \to 0 \tag{4.4}$$

in probability, we can redefine

$$\left[\hat{V}_n^0\right]^2 = \frac{2}{n(n-1)h^p} \sum_{i \neq j} \hat{\sigma}_n^{*2}(X_i) \hat{\sigma}_n^{*2}(X_j) K_h^2\left(X_i - X_j\right) \tag{4.5}$$

and the test statistic $T_n^0(\hat{\theta})$ accordingly. Since (4.4) and our assumptions imply $\hat{V}_n^0 - \tilde{V}_n^0 = o_P(1)$ uniformly in $h \in \mathcal{H}_n$, where here

$$\left[\tilde{V}_n^0\right]^2 = \frac{2}{n(n-1)h^p} \sum_{i \neq j} \sigma^{*2}(X_i) \sigma^{*2}(X_j) K_h^2\left(X_i - X_j\right), \tag{4.6}$$

the new test statistic $T_n^0(\hat{\theta})$ has the same standard normal asymptotic law under H_0 and potentially leads to a more powerful test. Lopez and Patilea (2006) provide sufficient conditions ensuring $\sup_{x \in \mathcal{X}} |\hat{\sigma}_n^{*2}(x) - \sigma_n^{*2}(x)| \to 0$, in probability, regardless of whether H_0 is true, where $\sigma_n^{*2}(\cdot)$ is defined like $\hat{\sigma}_n^{*2}(\cdot)$ but with estimated synthetic observations $\hat{Y}_i^*$ replaced with the true (unknown) ones Y_i^*. To obtain (4.4), their result can be completed by the arguments for i.i.d. data like in Horowitz and Spokoiny (2001) or Guerre and Lavergne (2005) allowing to deduce $\sup_{x \in \mathcal{X}} |\sigma_n^{*2}(x) - \sigma^{*2}(x)| \to 0$ in probability. In the WLS approach, the question of how to build an estimate of the variance of $nh^{p/2} Q_n^1(\hat{\theta})$ that (theoretically) performs better than $\hat{V}_n^1$ when H_0 is not true seems harder and therefore is left open. ■

Remark 2. The tests we propose depend on the choice of the smoothing parameter $h \in \mathcal{H}_n$. In section 5 we provide empirical evidence on the behavior of our tests with different bandwidths. On the other hand, following a well-known data-driven method for choosing the smoothing parameter, in the synthetic data approach we can define

$$T_n^{opt} = \max_{h \in \mathcal{H}_{1n}} T_n^0(\hat{\theta})$$

where the maximum is taken over a finite subset $\mathcal{H}_{1n} \subset \mathcal{H}_n$. Typically, $\mathcal{H}_{1n}$ is a geometric grid in $\mathcal{H}_n$ and the number of elements in $\mathcal{H}_{1n}$ increases as $n \to \infty$. See Horowitz and Spokoiny (2001). The resulting test is

$$\text{“Reject } H_0 \text{ when } T_n^{opt} \geq t_\alpha^{opt} \text{”}, \tag{4.7}$$

where t_α^{opt} is the $\alpha-$level critical value for T_n^{opt}. Like in the non-censored case, this critical value cannot be evaluated in applications because θ_0 and the law of the errors ε_i are unknown. Horowitz and Spokoiny (2001) proposed a simulation procedure for approximating the critical value t_α^{opt}. Their procedure can be adapted to our SD test when the test statistic $T_n^0(\hat{\theta})$ is defined using the standard deviation estimate $\hat{V}_n^0$ introduced by equation (4.5). The detailed investigation of this issue will be considered elsewhere. ■

4.3 Behavior of the tests under the alternatives

Consider a sequence of measurable functions $\lambda_n(x)$, $n \geq 1$, and the sequence of alternatives

$$H_{1n} : Y_{in} = f(\theta_0, X_i) + \lambda_n(X_i) + \varepsilon_i, \qquad 1 \leq i \leq n. \tag{4.8}$$

For simplicity, assume that there exists some constant M_λ such that for all $n \geq 1$, $0 \leq |\lambda_n(\cdot)| \leq M_\lambda < \infty$.

Assumption 8 *(i) The censoring times $C_1, ..., C_n$ represent an independent sample from the continuous distribution function G (the same for each n) and are independent of the variables $Y_{1n}, ..., Y_{nn}$ with continuous distribution function $F^{(n)}$.*

(ii) For each n, $\mathbb{P}(Y_{1n} \leq C_1 \mid X_1, Y_{1n}) = \mathbb{P}(Y_{1n} \leq C_1 \mid Y_{1n})$.

Notice that the second part of this assumption is always true if C is independent of ε and X. Now, for each n define $T_{in} = Y_{in} \wedge C_i$ and $\delta_{in} = \mathbf{1}_{\{Y_{in} \leq C_i\}}$, $i = 1, ..., n$, and let $H^{(n)}$ denote the distribution function of $T_{1n}, ..., T_{nn}$, that is $H^{(n)}(y) = \mathbb{P}(T_{1n} \leq y)$. Let us point out that the two test statistics we propose rely on the Kaplan-Meier estimator that is computed from the observations (T_{in}, δ_{in}), $i = 1, ..., n$. If $\lambda_n(\cdot)$ changes with n, the law of the observations is different for each n. Therefore, in order to control the jumps of the Kaplan-Meier estimator and the conditional variance of the residuals $U_i^\beta(\theta)$ we need the following assumption.

Assumption 9 *(i) There exist some constants c_{inf}, c_{sup} such that for each $x \in \mathcal{X}$*

$$0 < c_{inf} \leq \mathbb{E}\left[\varepsilon^2 \mid X = x\right] \leq \mathbb{E}\left[\left\{1+\varepsilon^2\right\}\{1 - G(Y_{1n})\}^{-1} \mid X = x\right] \leq c_{sup} < \infty.$$

(ii) There exists some constant M such that $\forall n \geq 1$, $\mathbb{E}\left[\{1+\varepsilon^4\}\,\gamma(Y_{1n})^4\right] \leq M < \infty$ where $\gamma(Y_{1n}) = \delta_{1n}\{1 - G(Y_{1n})\}^{-1}$.

(iii) Let $F^{(n)}_{Y|X=x}(y) = \mathbb{P}\left(Y_{1n} \leq y \mid X_1 = x\right)$ and

$$q^{(n)}_{\rho}(x) = \int \{|y| + 1\} C^{(n)}(y)^{1/2+\rho} dF^{(n)}_{Y|X=x}(y)$$

where

$$C^{(n)}(y) = \int_{-\infty}^{y} \frac{dG(t)}{[1 - H^{(n)}(t)][1 - G(t)]} \vee 1.$$

There exist $0 < \rho < 1/2$ and a function $q_\rho(x)$ with $\mathbb{E}[q^2_\rho(X)] < \infty$ such that for all n, $0 \leq q^{(n)}_\rho \leq q_\rho$.

Let $\hat{V}^{\beta}_n(\theta)^2$ be the estimator obtained after replacing $\hat{\theta}$ with θ on the right-hand side of (3.9). Once again, our purpose is to transfer the problem of consistency against the alternatives H_{1n} in the classical i.i.d. framework. The first step in this transfer is realized in a general setup in the following lemma proved in the Appendix. Next, we will be more specific on the type of alternatives considered in order to derive the asymptotic consistency.

Lemma 4.3 *Let Assumptions 4-(i) and (ii), 5, 6, 8 and 9-(ii) and (iii) hold true. Then, under the alternatives H_{1n}, for $\beta = 0$ or 1*

$$\left|Q^{\beta}_n(\theta) - \tilde{Q}^{\beta}_n(\theta)\right| \leq \left[\tilde{Q}^{\beta}_n(\theta) + R_{n1}\right]^{1/2} R^{1/2}_{n2} - R_{n3} + R_{n2} - R_{n4}$$

with $\sup_{\theta\in\Theta,\, h\in\mathcal{H}_n}\left\{h^p\left|R_{n1}\right| + \left|R_{n2}\right| + h^{p/2}\left|R_{n3}\right| + \left|R_{n4}\right|\right\} = O_P(n^{-1})$.

4.3.1 Consistency against a fixed alternative

Consider the alternative

$$H_1 : Y = m(X) + \varepsilon,$$

where $\mathbb{E}\left(\varepsilon \mid X\right) = 0$ a.s. and, for simplicity, we assume $0 \leq |m(\cdot)| \leq M_\lambda < \infty$ for some constant M_λ. The following assumption identifies the limit of $\hat{\theta}$ the SD or WLS estimator and states that the regression model is wrong.

Assumption 10 *There exists $\bar{\theta}$ an interior point of Θ such that*

$$\text{for any } \theta \in \Theta \setminus \{\bar{\theta}\}, \quad 0 < \mathbb{E}\left[\left\{m(X) - f\left(\bar{\theta}, X\right)\right\}^2\right] < \mathbb{E}\left[\left\{m(X) - f\left(\theta, X\right)\right\}^2\right].$$

Theorem 4.4 *Let Assumption 10, Assumption 9-(i) and the assumptions of Lemma 4.3 hold true. Under H_1, for $\beta = 0$ or 1*

$$\sup_{h \in \mathcal{H}_n} \left| Q_n^{\beta}(\hat{\theta}) - \mathbb{E}\left[\left\{m(X) - f\left(\bar{\theta}, X\right)\right\}^2 g(X)\right]\right| = o_P(1) \quad \text{and} \quad \sup_{h \in \mathcal{H}_n} |\hat{V}_n^{\beta} - c| = o_P(1),$$

where $c > 0$ is some constant. Consequently, the tests in (3.8) are consistent.

See the Appendix for the proof. It is worthwhile to notice that the limit of $Q_n^{\beta}(\hat{\theta})$ under the alternative H_1 does not depend on the censoring and is the same for $\beta = 0$ or $\beta = 1$. However, the limits of the standard deviations $\hat{V}_n^{\beta}$ depend on β and the degree of censoring in the data (see Lemma A.8). In general, our tests lose power if the degree of censoring increases. Moreover, looking at the limits of $\hat{V}_n^{\beta}$ for $\beta = 0$ and $\beta = 1$, one notices that none of the two tests is more powerful than the other, that means depending on the law of (Y, C), either the SD or WLS test will perform better.

4.3.2 Consistency against Pitman local alternatives

Let $\lambda(\cdot)$ be a measurable function of X and consider the sequence of alternatives

$$H_{1n}: \; Y_{in} = f\left(\theta_0, X_i\right) + r_n \lambda(X_i) + \varepsilon_i, \qquad 1 \leq i \leq n,$$

with $r_n \downarrow 0$ when $n \to \infty$. For simplicity, we will assume that

$$\lambda(\cdot) \text{ is a bounded function and } \mathbb{E}\left[\lambda(X) \nabla_{\theta} f\left(\theta_0, X\right)\right] = 0. \tag{4.9}$$

The latter condition will make $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$. See Lemma A.8. The following result, proved in the Appendix, implies that our tests are consistent against the local alternatives H_{1n}, if r_n decreases slower than $n^{-1/2} h^{-p/4}$.

Theorem 4.5 *Let Assumption 9-(i), the assumptions of Lemma 4.3 and condition (4.9) hold true. Under H_{1n}, for $\beta = 0$ or 1 the test statistics $T_n^{\beta}(\hat{\theta})$ converge in law to a normal distribution $N(\mu, 1)$ with $\mu > 0$, provided that $r_n = n^{-1/2} h^{-p/4}$.*

4.3.3 Consistency against a sequence of smooth alternatives

Here, we provide conditions under which our tests are consistent against alternatives H_{1n} like in (4.8) defined by functions $\lambda_n(\cdot)$ in a Hölder smoothness class that vanish as $n \uparrow \infty$. The regularity s of the Hölder class is supposed *known* and the rate to which the functions $\lambda_n(\cdot)$ approach zero can be made arbitrarily close to the optimal rate of testing $n^{-2s/(4s+p)}$, when $s > 5p/4$. We have to be more restrictive on the regularity s (the usual condition being $s \geq p/4$, see Horowitz and Spokoiny, 2001) because of our conditions on the left endpoint of the bandwidth range $\mathcal{H}_n$. See Assumption 6-(ii) and the subsequent comments. For $L > 0$, define the Hölder class $C(L, s)$ as

$$C(L,s) = \{f(\cdot) : |f(x_1) - f(x_2)| \leq L|x_1 - x_2|^s, \, \forall x_1, x_2 \in \mathcal{X}\}, \quad \text{for } s \in (0,1],$$

while for $s > 1$, $C(L, s)$ is the class of functions having the $[s]$-th partial derivatives in $C(L, s - [s])$, where $[s]$ denotes the integer part of s. As a corollary of the following theorem, the optimal rate of testing parametric mean-regressions when s is known is not altered by the censorship, provided that $s > 5p/4$. The proof of the theorem is postponed to the Appendix.

Theorem 4.6 *Let Assumption 9-(i) and the assumptions of Lemma 4.3 hold. Moreover, the density $g(\cdot)$ is bounded from below by a positive constant. Let κ_n, $n \geq 1$ be a sequence of positive real numbers. Consider a sequence of functions $\lambda_n(\cdot)$ such that for all $n \geq 1$, $\lambda_n(\cdot) \in C(L, s)$ for some known $s > 5p/4$ and some $L > 0$. Moreover, $\mathbb{E}\left[\lambda_n^2(X)\right] \to 0$ as $n \to \infty$ and for each $n \geq 1$, $\mathbb{E}[\lambda_n(X) \nabla_\theta f(\theta_0, X)] = 0$ and*

$$\|\lambda_n\|_n := \left[n^{-1} \sum_{i=1}^{n} \lambda_n^2(X_i) \right]^{1/2} \geq \kappa_n n^{-\frac{2s}{4s+p}}. \tag{4.10}$$

If h is of order $n^{-2/(4s+p)}$, the tests defined in (3.8) are consistent against the alternatives H_{1n} defined by the functions $\lambda_n(\cdot)$ whenever κ_n diverges.

Remark 2 *(continued).* In Theorem 4.6 we supposed that the regularity s is known and thus the rate of the bandwidth that allows to detect departures from the null hypothesis like in (4.10) is known. More generally, it would be useful to have a data-driven selection procedure for h that adapts to the unknown smoothness of the functions $\lambda_n(\cdot)$ and that allows these functions to converge to zero at a rate which is arbitrarily close to the fastest possible rate. In the case of non-censored responses, if s is unknown but $s \geq p/4$, the optimal rate of testing is $(n^{-1}\sqrt{\log\log n})^{2s/(4s+p)}$, see for instance Horowitz and Spokoiny (2001). The maximum test procedure (4.7) represents a potential solution in the synthetic data testing approach. Consider the test statistic built with the true synthetic observations and the estimate of the parameter θ_0, that is $\tilde{T}_n^0(\hat{\theta}) = nh^{p/2}\tilde{Q}_n^0(\hat{\theta})/\tilde{V}_n^0$ with $\tilde{V}_n^0$ defined like in (4.6). Suppose that under the alternatives H_{1n} defined by functions $\lambda_n(\cdot)$ like in Theorem 4.6 with some suitable $\kappa_n \uparrow \infty$, we have

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{h\in\mathcal{H}_{1n}} \tilde{T}_n^0(\hat{\theta}) \geq t_\alpha^b \right) = 1, \tag{4.11}$$

where t_α^b is some suitable critical value. Then, by Lemma 4.3, it is expected that $\mathbb{P}\left(\max_{h\in\mathcal{H}_{1n}} T_n^0(\hat{\theta}) \geq t_\alpha^b\right) \to 1$. In view of the proof of Theorem 4.6, we argue that any κ_n such that $\kappa_n \left[\log\log n\right]^{-s/(4s+p)} \to \infty$ ensures condition (4.11) when $\mathcal{H}_{1n}$ is a geometric grid like in Horowitz and Spokoiny (2001). The detailed investigation of these issues will be considered elsewhere. ■

5 Empirical studies

To investigate the finite sample properties of our tests and to compare them to the alternative tests of Stute *et al.* (2000), we conducted several simulation and real data experiments. The results are presented below.

5.1 Simulation experiments

The regression model considered in simulations was $Y = \theta_{01} + \theta_{02}X + \varepsilon$ with X uniformly distributed on the interval $[-\sqrt{3}, \sqrt{3}]$ and ε a standard normal residual term. A linear regression function appears, for instance, in the so-called accelerated failure time (AFT) model that has found considerable interest in the survival data literature. The true parameters are $(\theta_{01}, \theta_{02}) = (1, 3)$ and C has an exponential distribution of mean μ. The parameter μ served to control the proportion of censored observations that was fixed to 30%, 40% or 50%.

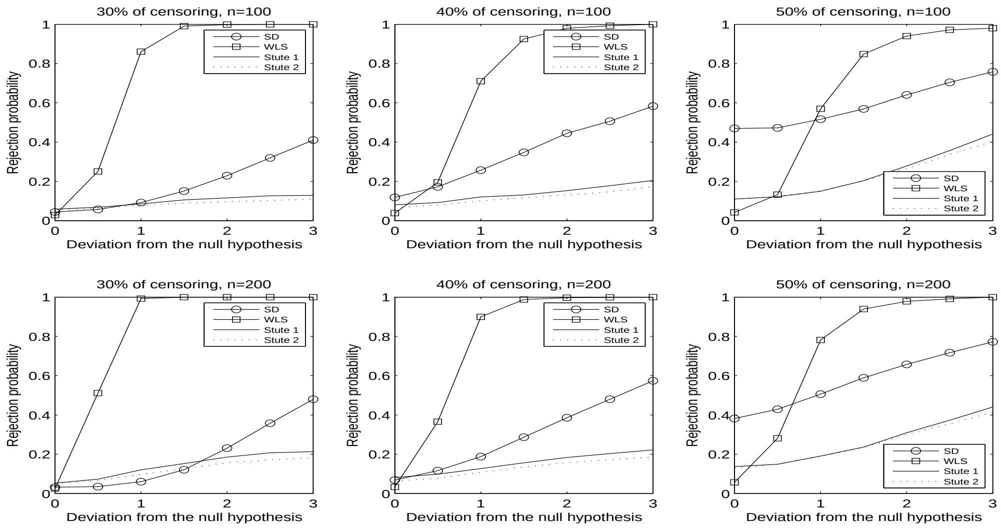

Figure 1: Rejection probabilities for T_n^{SD}, T_n^{WLS}, D_n (Stute 1) and W_n^2 (Stute 2) test statistics with cosine alternatives.

First, the linear regression model was tested against alternatives with the form

$$H_1: \; Y_i = \theta_{01} + \theta_{02} X_i + d \cos(2\pi (X_i/\sqrt{3})) + \varepsilon_i, \qquad 1 \leq i \leq n,$$

with $d \in \{0.5, 1, ..., 2.5, 3\}$. The way the alternatives were defined rendered the amount of censoring practically stable on the null and under the alternatives. The levels considered were $\alpha = 0.05$ and $\alpha = 0.1$. We took $n = 100$ and $n = 200$ and for each sample size we generated 5000 samples. We used a gaussian kernel and the bandwidth $h = 0.1$ for the kernel-based tests. The test statistic T_n^{SD} (resp. T_n^{WLS}) was built using the estimator $\hat{\theta}^{SD}$ (resp. $\hat{\theta}^{WLS}$). The critical values for our tests were those given by the standard normal law while for the tests proposed by Stute *et al.* (2000) we followed their bootstrap procedure (with 5000 bootstrap samples). The asymptotic distribution of test statistics D_n and W_n^2 used by Stute *et al.* (2000) depend on the asymptotic distribution of the estimator of θ_0. To focus the attention on the performances of the testing approaches, we computed the values of D_n and W_n^2 using the true values of the parameters θ_{01}, θ_{02}. This resulted in improved rejection probabilities under the null and under the alternatives for the corresponding tests. The results of the simulations are presented in Figure 1. To save space, only the results for $\alpha = 0.05$ are reported, the case $\alpha = 0.1$ being very similar. This first empirical investigation shows that in the setup considered, the test based on T_n^{WLS}

outperforms the test built with T_n^{SD} and the tests obtained with the weighted marked empirical process approach of Stute *et al.* (2000). The level of the WLS kernel-based test is satisfactory close to the nominal level for all probabilities of censoring considered. On contrary, the level of the SD-based test drastically deteriorates when the probability of censoring increases. With a few minor exceptions, the rejection probabilities under the alternatives are higher or much higher for the kernel-based tests than for the tests based on the marked empirical process approach.

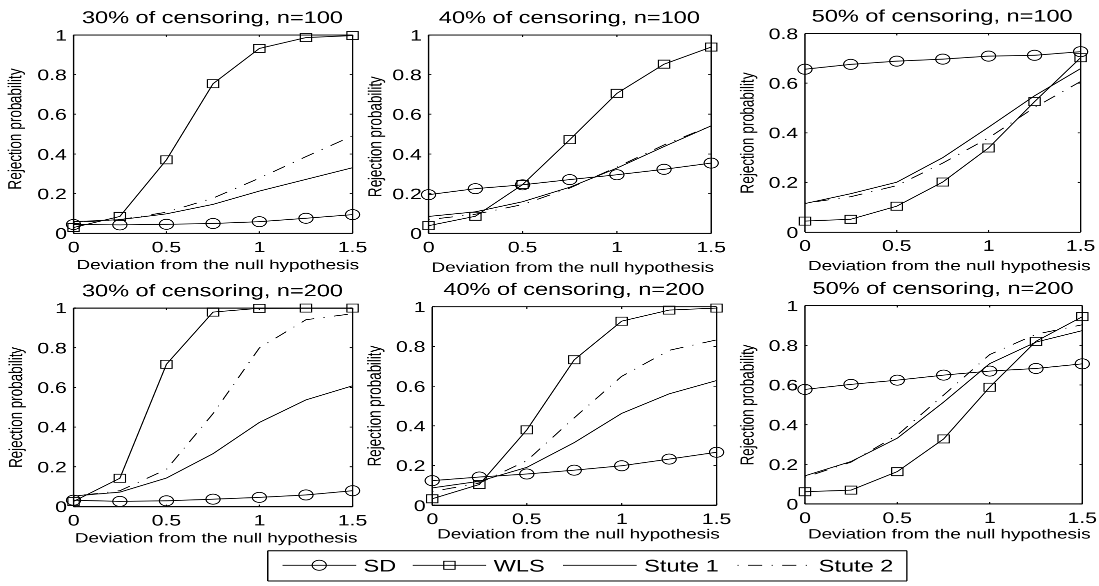

Figure 2: Rejection probabilities for T_n^{SD}, T_n^{WLS}, D_n (Stute 1) and W_n^2 (Stute 2) test statistics with quadratic alternatives.

The literature on nonparametric models checks contains evidence that sine and cosine alternatives are easily detected by smoothing based procedures. To provide a fair comparison between the alternative approaches, we considered a second simulation experiment where the same linear regression model was tested against the alternatives

$$H_1 : \; Y_i = \theta_{01} + \theta_{02} X_i + d(X_i^2 - 1) + \varepsilon_i, \qquad 1 \le i \le n, \tag{5.1}$$

with $d \in \{0.25, 0.5, ..., 1.25, 1.5\}$. The level was $\alpha = 0.05$. We took the same sample sizes ($n = 100$ and $n = 200$) and 5000 replications for each sample size. The bandwidth was $h = 0.1$. The test statistics and the critical values were calculated as in the first example. The results of this second experiment are presented in Figure 2. The performances of empirical process based tests are now always better than those of SD-based test. The WLS kernel-based test is still the best procedure when 30% or 40% of responses are censored. The tests of Stute *et al.* (2000) have slightly better power when half of the lifetimes Y are censored, but their rejection probability under the null hypothesis is less satisfactory. Meanwhile, the standard normal critical values are still satisfactory for our WLS test.

As pointed out by a referee, it is important to have some insight on the performances of the kernel-based tests when the bandwidth h changes. To investigate this issue, we considered the same linear regression model and sample sizes as before and a quadratic alternative like in (5.1) with $d = 1$. For each sample size 5000 replications were used. The bandwidths selected to compute T_n^{WLS} and T_n^{SD} were $h \in \{0.025, 0.05, ..., 0.325, 0.35\}$. These bandwidth values are quite common for smoothing with samples like those generated here. The results obtained with T_n^{WLS} are depicted in Figure 3. One could notice the almost stable rejection probabilities under the null and under the alternative for a wide range of bandwidths. We obtained a similar picture (not reported herein) confirming the failure of the SD-based test for the whole range of bandwidths considered. These results provide useful guidance for the applications.

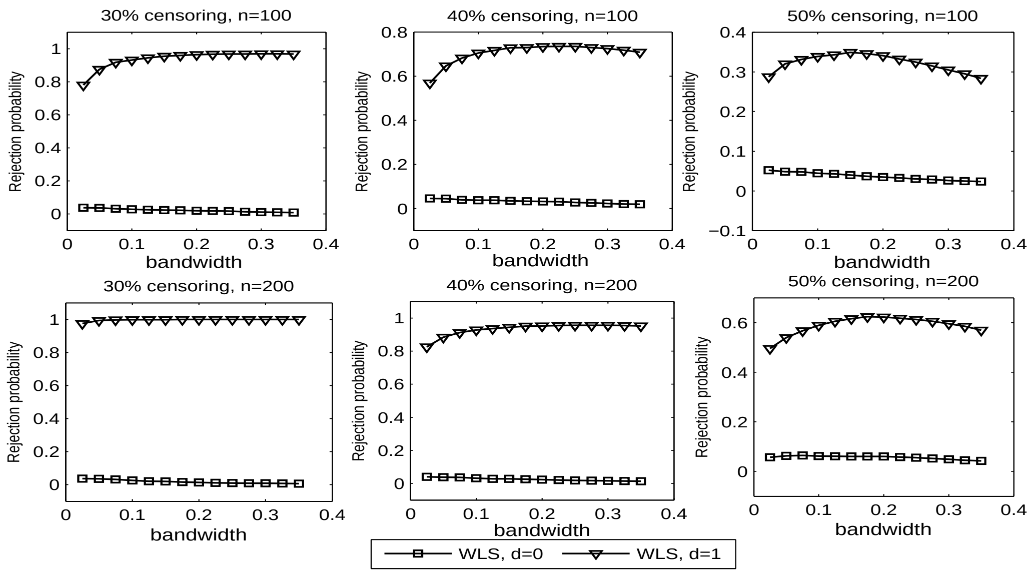

Figure 3: Rejection probabilities for WLS kernel-based test under the null and under a quadratic alternative when the bandwidth h varies.

Finally, in view of the poor performances of the SD-based test, one may want to use the bootstrap for calibrating the critical values. When the response Y is not censored, a classical bootstrap procedure consists in drawing n i.i.d. random variables ω_i independent from the original sample with $\mathbb{E}(\omega_i) = 0$, $\mathbb{E}(\omega_i^2) = 1$, and $\mathbb{E}(\omega_i^4) < \infty$, and to generate bootstrap observations of Y as $Y_i^{(b)} = f(\hat{\theta}, X_i) + \hat{\tau}_n(X_i)\omega_i$, $i = 1, ..., n$. Here, $\hat{\tau}_n^2(\cdot)$ is a non parametric estimator of the conditional variance of Y_i given X_i. A bootstrap test statistic is built from the bootstrap sample as was the original test statistic. When this scheme is repeated many times, the empirical $(1-\alpha)$th quantile of the bootstrapped test statistics gives the bootstrap critical value. This critical value is then compared to the initial test statistic. See, for instance, Guerre and Lavergne (2005).

Table 1: Rejection probabilities with standard normal critical values (*WLS* and *SD* columns) and bootstrap critical values (*SD bootstrap* column) – quadratic alternative

n	deviation	censoring	SD	SD bootstrap	WLS
100	$d = 0$	40%	0.168	0.07	0.039
		50%	0.661	0.223	0.055
	$d = 0.75$	40%	0.242	0.163	0.486
		50%	0.696	0.401	0.225
	$d = 1.5$	40%	0.31	0.277	0.932
		50%	0.726	0.584	0.703
200	$d = 0$	40%	0.126	0.063	0.045
		50%	0.554	0.128	0.051
	$d = 0.75$	40%	0.161	0.115	0.728
		50%	0.643	0.333	0.343
	$d = 1.5$	40%	0.268	0.249	0.998
		50%	0.701	0.565	0.939

When Y is censored, by property (2.3), $\mathbb{E}[Y \mid X]$ (resp. $\mathbb{E}[Y^2 \mid X]$) is equal to $\mathbb{E}[\delta\{1 - G(T-)\}^{-1}T \mid X]$ (resp. $\mathbb{E}[\delta\{1 - G(T-)\}^{-1}T^2 \mid X]$) and thus the conditional variance of Y can still be estimated from data. The additional difficulty with censored data is that one also needs bootstrap samples for the censoring times C_i in order to build bootstrap samples for $T_i = Y_i \wedge C_i$ and $\delta_i = \mathbf{1}_{\{Y_i \le C_i\}}$. Bootstrap censoring times could be generated, for instance, using the Kaplan-Meier estimator of G. With at hand the bootstrap observations $T_i^{(b)}$ and $\delta_i^{(b)}$, one could follow the classical bootstrap methodology and compute bootstrap critical values for the T_n^{SD} test statistic. The study of the asymptotic validity of this procedure in the presence of censoring will be undertaken elsewhere. Here, we investigate the empirical properties of this bootstrap procedure when the alternatives (5.1) are considered. For simplicity, the conditional variance of Y is supposed to be known. The number of replications was 1000 and for each replication 399 bootstrap samples were generated. We used the bandwidths $h \in \{0.05, 0.1, 0.15, 0.2, 0.25\}$. The results are presented in Table 1 for the case where 40% and 50% of the responses were censored, $h = 0.1$ and $\alpha = 0.05$. The results for the other bandwidths were quite similar. Let us notice that the bootstrap critical values improve the rejection probability of the SD-based test under the null hypothesis. However, the WLS kernel-based test, applied with the standard normal critical values, is still the best procedure.

5.2 Real data application

We now illustrate our test procedures using data from the Stanford Heart Transplant program between October 1967 and February 1980. During this period, 184 of the 249 patients admitted to the program received a heart transplantation. Patients alive beyond February 1980 were considered censored. For purposes of comparison with the empirical investigations of Stute *et al.* (2000), Miller and Halpern (1982) and Wei *et al.* (1990), we concentrate our analysis on the subsample of 152 patients who had complete tissue typing and survived at least 10 days. Among the 152 cases, 55 were censored, that is 36.18%. The parametric regression model tested is the linear regression for $\log_{10}$ of time to death versus age and age squared. The covariates were standardized and three values were used for the bandwidth h (0.15, 0.2 and 0.25). We also used three different bandwidths (0.18, 0.36 and 0.54) for the nonparametric estimate of the conditional variance of the response that is needed to generate bootstrap samples. Here, only the results corresponding to the value 0.36 are presented, the other results being similar. The kernel was gaussian and 399 bootstrap samples were used for calibrating the SD-based test. The $p-$values are reported in Table 2.

We see that the $p-$value of the SD-based test obtained with the bootstrap is much larger than the $p-$value obtained with standard normal asymptotic approximation. Wei *et al.* (1990) and Stute *et al.* (2000) came to the conclusion that the linear model that we test here cannot be rejected (the $p-$value obtained by Wei *et al.* was 0.67, while the $p-$values of D_n and W_n^2 statistics of Stute *et al.* were 0.8413 and 0.8793, respectively). Our results confirm this conclusion.

Table 2: $P-$values of the SD, SD bootstrap and WLS tests with Stanford Heart Transplant Data

Test	$h = 0.15$	$h = 0.2$	$h = 0.25$
SD	0.03	0.03	0.027
WLS	0.652	0.748	0.798
SD bootstrap	0.185	0.198	0.228

Appendix

First, we prove some technical lemmas. We refer to Nolan and Pollard (1987) for the definition of Euclidean classes of functions. Below, M, c, c_1, ... are constants that may be different from line to line.

A.1 Technical lemmas

The point (ii) of the following lemma provides a bound for the difference between the weights W_{in} and the ideal weights one would obtain if G were known. Here, for each sample size n, the lifetimes Y are supposed independent with a same law which may depend on n. This generality is needed under alternatives changing with the sample size.

Lemma A.1 *Let $Y_{1n}, ..., Y_{nn}$ be an independent sample from a continuous distribution function $F^{(n)}$, $n \geq 1$. Independent of these, let $C_1, ..., C_n$ be an independent sample from a continuous distribution function G (the same for each n). Let $T_{in} = Y_{in} \wedge C_i$ and*

$\delta_{in} = \mathbf{1}_{\{Y_{in} \le C_i\}}$, $i = 1, ..., n$, *and for each* n, *let* $H^{(n)}$ *denote the distribution function of* $T_{1n}, ..., T_{nn}$. *Denote* $\gamma(T_{in}) = \delta_{in}[1 - G(T_{in})]^{-1}$ *and let* $T_{(n)n} = \max_{1 \le i \le n} T_{in}$. *Then,*

i)

$$\sup_{1 \le i \le n} \frac{1 - \hat{G}(T_{in}-)}{1 - G(T_{in})} = O_P(1) \quad \text{and} \quad \sup_{1 \le i \le n} \frac{1 - G(T_{in})}{1 - \hat{G}(T_{in}-)} = O_P(1); \tag{A.1}$$

ii) Under Assumption 9, for all $0 \le \alpha \le 1/2$ *and* $\eta > 0$,

$$|nW_{in} - \gamma(T_{in})| \le \frac{\delta_{in}}{1 - G(T_{in})} \{C^{(n)}(T_{in})\}^{\alpha+\eta} \times O_P(n^{-\alpha}),$$

where the $O_P(n^{-\alpha})$ *factor does not depend on* i.

Proof. For the sake of simplicity, we only consider the case $Y_{in} = Y_i$ and $T_{in} = T_i$. The general proof can be found in Lopez and Patilea (2006).

i) Since by assumption $P(Y_i = C_i) = 0$, we can redefine $1 - \delta_i = \mathbf{1}_{\{C_i \le Y_i\}}$ and study $\hat{G}$ as the Kaplan-Meier estimator of the lifetimes C_i in presence of the censoring times Y_i. The first part of (A.1) follows from Theorem 3.2.4 in Fleming and Harrington (1991). The second part follows for instance as a consequence of Theorem 2.2 in Zhou (1991).

ii) Fix $\eta > 0$ arbitrarily. Since $\int_a^{\tau_H} C^{-1-2\eta}(y)dC(y) < \infty$, for some $a > 0$, apply Theorem 1 in Gill (1983) to see that

$$\sup_{y \le T_{(n)}} [C(y)]^{-1/2-\eta} |Z(y)| = O_P(1), \tag{A.2}$$

where $Z = \sqrt{n}\{\hat{G} - G\}\{1 - G\}^{-1}$ is the Kaplan-Meier process. Next, the proof can be completed by using the definitions of W_{in} and $\gamma(\cdot)$ and elementary algebra. ∎

Let A_h be the $n \times n$ symmetric matrix with generic element

$$a_{ij}(h) = [h^p n(n-1)]^{-1} K_h(X_i - X_j)\mathbf{1}_{\{i \ne j\}}. \tag{A.3}$$

Lemma A.2 *Let* $v_1, ..., v_n$ *and* $w_1, ..., w_n$ *be sequences of real numbers. Suppose that Assumptions 4 (i)-(ii) and 6 (ii) hold true. If*

$$U(h) = \frac{1}{n^2 h^p} \sum_{1 \le i \ne j \le n} v_i w_j K_h(X_i - X_j),$$

then

$$\sup_{h \in \mathcal{H}_n} |U(h)| \le O_P(1) \left[\frac{1}{n}\sum_{i=1}^n v_i^2\right]^{1/2} \left[\frac{1}{n}\sum_{i=1}^n w_i^2\right]^{1/2}.$$

For the proof of this result, recall that for any $n-$dimensional vectors z_1, z_2, $|z_1' A_h z_2| \leq \|A_h\|_2 \|z_1\| \|z_2\|$. Guerre and Lavergne (2005) proved that $\|A_h\|_2 = O_P(n^{-1})$ under the assumptions of Lemma A.2, while Lopez and Patilea (2006) showed that this order in probability holds uniformly in $h \in \mathcal{H}_n$. These facts prove Lemma A.2.

Lemma A.3 *Let $X_1, X_2, ...$ be a sample as in Assumption 4-(i) and (ii) and let Assumption 6 hold true. For each $n \geq 1$, let $u_{1n}, ..., u_{nn}$ be a sequence of random variables that are independent given $X_1, ..., X_n$. For each n and i, the law of u_{in} given $X_1, ..., X_n$ depends only on X_i. Assume $\mathbb{E}(u_{in} \mid X_i) = 0$ and $\mathbb{E}(u_{in}^2 \mid X_i) = \sigma_n^2(X_i)$ and suppose that for each x and n we have $0 \leq \sigma_n^2(x) \leq \overline{\sigma}_n^2 < \infty$. Then*

$$\frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} u_{in} u_{jn} \frac{1}{h^p} K_h(X_i - X_j) = \overline{\sigma}_n^2 O_P\left(n^{-1} h^{-p/2}\right). \tag{A.4}$$

Let $\lambda_n(\cdot)$, $n \geq 1$ be a sequence of measurable functions and let

$$U_n = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \lambda_n(X_i) u_{jn} \frac{1}{h^p} K_h(X_i - X_j).$$

If A_h is defined as in (A.3) and $\|\lambda_n\|_n^2$ denotes $n^{-1} \sum_{i=1}^n \lambda_n^2(X_i)$, then

$$\mathbb{E}[|U_n| \mid X_1, ..., X_n] \leq c\, \overline{\sigma}_n n^{1/2} \|A_h\|_2 \|\lambda_n\|_n$$

for some finite constant c independent of n and of the sequence $\lambda_n(\cdot)$, $n \geq 1$.

Proof. By elementary calculus, the variance of the degenerate $U-$statistic in (A.4) is of order $n^{-2} h^{-p}$ and thus we obtain stated rate from Chebyshev's inequality. Next, following Guerre and Lavergne (2005, Lemma 3), let

$$\overline{\lambda}_n(X_i) = \frac{1}{n(n-1)} \sum_{j=1, i \neq j}^{n} \lambda_n(X_j) \frac{1}{h^p} K_h(X_i - X_j).$$

By Marcinkiewicz-Zygmund inequality and Jensen inequality

$$\begin{aligned}
\mathbb{E}\left[\left|\sum_{i=1}^n u_{in} \overline{\lambda}_n(X_i)\right| \mid X_1, ..., X_n\right] &\leq c\,\mathbb{E}\left[\left(\sum_{i=1}^n u_{in}^2 \overline{\lambda}_n^2(X_i)\right)^{1/2} \mid X_1, ..., X_n\right] \\
&\leq c\left[\sum_{i=1}^n \mathbb{E}\left(u_{in}^2 \mid X_i\right) \overline{\lambda}_n^2(X_i)\right]^{1/2} \leq c\, \overline{\sigma}_n \left[\sum_{i=1}^n \overline{\lambda}_n^2(X_i)\right]^{1/2} \leq c\, \overline{\sigma}_n n^{1/2} \|A_h\|_2 \|\lambda_n\|_n,
\end{aligned}$$

where c is a constant independent of n and of the sequence $\lambda_n(\cdot)$, $n \geq 1$. ■

A.2 Proofs

This section starts with several lemmas that will be used in the proof of Theorem 4.1.

Lemma A.4 *Let the assumptions of Theorem 4.1 hold and fix* $\zeta \in (0, 1/2)$ *arbitrarily. Under* H_0*, for* $\beta = 0$ *or 1,* $\sup_{h\in\mathcal{H}_n} h^{\zeta} \left| Q_n^{\beta}(\hat\theta) - Q_n^{\beta}(\theta_0) \right| = O_P(n^{-1})$.

Proof. By definition $\hat U_i^{\beta}(\hat\theta) - \hat U_i^{\beta}(\theta_0) = (nW_{in})^{\beta}[f(\hat\theta, X_i) - f(\theta_0, X_i)]$, where by convention $(nW_{in})^{\beta} = 1$ for $\beta = 0$ and $(nW_{in})^{\beta} = nW_{in}$ for $\beta = 1$. A similar convention applies for $\gamma^{\beta}(T_i)$. Write

$$
\begin{aligned}
Q_n^{\beta}(\hat\theta) &= Q_n^{\beta}(\theta_0) + 2\sum_{i\neq j} \hat U_i^{\beta}(\theta_0)(nW_{jn})^{\beta}[f(\hat\theta, X_j) - f(\theta_0, X_j)]a_{ij}(h) \\
&\quad + \sum_{i\neq j}\left(n^2 W_{in}W_{jn}\right)^{\beta}[f(\hat\theta, X_i) - f(\theta_0, X_i)][f(\hat\theta, X_j) - f(\theta_0, X_j)]a_{ij}(h) \\
&= Q_n^{\beta}(\theta_0) + 2Q_{n1}^{\beta}(\hat\theta, \theta_0) + Q_{n2}^{\beta}(\hat\theta, \theta_0).
\end{aligned}
$$

By Assumption 5, there exists some constant c independent of h such that

$$
\begin{aligned}
\left| Q_{n2}^{\beta}(\hat\theta, \theta_0) \right| &\leq c\|\hat\theta - \theta_0\|^2 \times \sum_{i\neq j}(nW_{in})^{\beta}(nW_{jn})^{\beta} a_{ij}(h) = O_P(n^{-1}) \\
&\leq O_P(1)\,\|\hat\theta - \theta_0\|^2 \sum_{i\neq j}\gamma^{\beta}(T_i)\gamma^{\beta}(T_j)a_{ij}(h),
\end{aligned}
$$

where for the second inequality we used the first part of equation (A.1). As $\mathbb{E}[\gamma^2(T)] < \infty$ (by Assumption 4-(iv)) and $\hat\theta - \theta_0 = O_P(n^{-1/2})$ (see Delecroix *et al.* 2006), Lemma A.2 implies $\sup_{h\in\mathcal{H}_n}\left| Q_{n2}^{\beta}\left(\hat\theta, \theta_0\right)\right| = O_P\left(n^{-1}\right)$.

To investigate Q_{n1}^{β}, let

$$
\tilde Q_{n1}^{\beta}(\hat\theta, \theta_0) = \sum_{i\neq j} U_i^{\beta}(\theta_0)\gamma^{\beta}(T_j)[f(\hat\theta, X_j) - f(\theta_0, X_j)]a_{ij}(h).
$$

By Taylor expansion, Assumption 5(i), Lemma A.2 and $\mathbb{E}[U_i^{\beta}(\theta_0)^2 + \gamma^{\beta}(T)^2] < \infty$,

$$
\begin{aligned}
\tilde Q_{n1}^{\beta}\left(\hat\theta, \theta_0\right) &= \frac{(\hat\theta - \theta_0)'}{n(n-1)h^p}\sum_{i\neq j}\Big\{U_i^{\beta}(\theta_0)\gamma^{\beta}(T_j) \\
&\qquad \times \nabla_{\theta} f(\theta_0, X_j)K_h(X_i - X_j)\Big\} + \|\hat\theta - \theta_0\|^2 O_P(1) \\
&= h^{-p}(\hat\theta - \theta_0)'\tilde S_{n1}^{\beta}(h) + \|\hat\theta - \theta_0\|^2 O_P(1),
\end{aligned}
$$

with the $O_P(1)$ factor independent of h. For the zero mean $U-$process $\tilde{S}^{\beta}_{n1}(h)$ apply the Hoeffding decomposition and write it as a sum of degenerate $U-$processes of order 2 and 1, say $\tilde{S}^{\beta}_{n11}(h)$ and $\tilde{S}^{\beta}_{n12}(h)$, indexed by families defined by h that are Euclidean for square integrable envelopes (this property is ensured by the bounded variation of the kernel $\tilde{K}$, Lemma 22-(ii) of Nolan and Pollard 1987, and Lemma 5 of Sherman 1994). By Corollary 4 of Sherman (1994), the rate of the uniform convergence of $\tilde{S}^{\beta}_{n11}(h)$ is $O_P(n^{-1})$. Deduce $\sup_{h\in\mathcal{H}_n} h^{-p}|\tilde{S}^{\beta}_{n11}(h)| = O_P(n^{-1/2})$. On the other hand, $h^{-p}\tilde{S}^{\beta}_{n12}(h)$ writes like $n^{-1}\sum_{i=1}^{n} U^{\beta}_i(\theta_0)\phi_i$ with

$$\phi_i = \mathbb{E}[\gamma^{\beta}(T_j)\nabla_\theta f(\theta_0, X_j)h^{-p}K_h(X_i - X_j) \mid X_i].$$

Notice that $|\phi_i| \leq M$, for some constant M. Let $h_L \leq h_{min} \leq h_{L-1} < ... < h_1 < h_0 = h_{max}$ a grid of bandwidths with $h_l = h_{l-1}h^c_{max}$, $1 \leq l \leq L$, and $c > 0$ to be chosen below. By definition $\mathcal{H}_n \subset \bigcup_{l=1}^{L} H_l$, where $H_l = [h_l, h_{l-1}]$. Fix arbitrarily $\alpha \in (0,1)$ such that $1 - \zeta/p < \alpha$. For each $l = 1, ..., L$, by the definition of H_l and Sherman's (1994) Main Corollary

$$\begin{aligned}
\mathbb{E}\left[\sup_{h\in H_l} |n^{1/2}h^{\zeta-p}\tilde{S}^{\beta}_{n12}(h)|\right] &\leq h_l^{\zeta-p}\mathbb{E}\left[\sup_{h\in H_l} |n^{1/2}\tilde{S}^{\beta}_{n12}(h)|\right] \\
&\leq \Lambda_1 h_l^{\zeta-p}\left[\mathbb{E}\sup_{h\in H_l}\{h^{2p}\frac{1}{2n}\sum_{i=1}^{2n} U^{\beta}_i(\theta_0)^2\phi_i^2\}^{\alpha}\right]^{1/2} \\
&\leq \Lambda_2 h_l^{\zeta-(1-\alpha)p}\left(\frac{h_{l-1}}{h_l}\right)^{\alpha p}\left[\frac{1}{2n}\sum_{i=1}^{2n} U^{\beta}_i(\theta_0)^2\right]^{\alpha/2} \\
&= h^{a_l}_{max}\, O_P(1),
\end{aligned}$$

where Λ_1, Λ_2 are constants that depend on α and τ (and p) but not on n and l and $a_l = 1 + \{l\,[\zeta - (1-\alpha)\,p] - p\alpha\}\,c$. The Euclidean property for a square integrable envelope required in Sherman's Main Corollary is ensured by the bounded variation of the kernel $\tilde{K}$, Lemma 22-(ii) of Nolan and Pollard (1987) and Lemma 5 of Sherman (1994). Take c such that $1 + (\zeta - p)\,c > 0$. Looking at the sum of the geometric series with common ratio $h^{[\zeta-(1-\alpha)p]c}_{max}$ and starting term $h^{1+(\zeta-p)c}_{max}$, deduce that $\mathbb{E}\left[\sup_{h\in\mathcal{H}_n} |n^{1/2}h^{\zeta-p}\tilde{S}^{\beta}_{n12}(h)|\right] \to 0$. This and Chebyshev's inequality provide the order of $h^{\zeta-p}\tilde{S}^{\beta}_{n12}(h)$ uniformly in $h \in \mathcal{H}_n$. Collecting results and using $\|\hat{\theta} - \theta_0\| h^{-p}_{min} = o_P(1)$,

$$\sup_{h\in\mathcal{H}_n} h^{\zeta}\left|\tilde{Q}^{\beta}_{n1}\left(\hat{\theta}, \theta_0\right)\right| = O_P(n^{-1}).$$

Next, rewrite

$$
\begin{aligned}
Q_{n1}^{\beta}(\hat{\theta},\theta_0) &= \tilde{Q}_{n1}^{\beta}\left(\hat{\theta},\theta_0\right) \\
&\quad+\sum_{i\neq j}[\hat{U}_i^{\beta}(\theta_0)-U_i^{\beta}(\theta_0)]\gamma^{\beta}(T_j)\,[f(\hat{\theta},X_j)-f(\theta_0,X_j)]a_{ij}(h) \\
&\quad+\sum_{i\neq j}U_i^{\beta}(\theta_0)\left[(nW_{jn})^{\beta}-\gamma^{\beta}(T_j)\right][f(\hat{\theta},X_j)-f(\theta_0,X_j)]a_{ij}(h) \\
&\quad+\sum_{i\neq j}[\hat{U}_i^{\beta}(\theta_0)-U_i^{\beta}(\theta_0)]\left[(nW_{jn})^{\beta}-\gamma^{\beta}(T_j)\right][f(\hat{\theta},X_j)-f(\theta_0,X_j)]a_{ij}(h) \\
&= \tilde{Q}_{n1}^{\beta}\left(\hat{\theta},\theta_0\right)+\tilde{Q}_{n11}^{\beta}+\tilde{Q}_{n12}^{\beta}+\tilde{Q}_{n13}^{\beta}.
\end{aligned}
$$

To show the negligibility of $\tilde{Q}_{n11}^{\beta}$ to $\tilde{Q}_{n13}^{\beta}$ we can no longer use the quick argument of Lemma A.2 because the random variables we have to manipulate are no longer square integrable. Indeed, by definition

$$
\hat{U}_i^{\beta}(\theta_0)-U_i^{\beta}(\theta_0)=[nW_{in}-\gamma(T_i)]\,[T_i-\beta f(\theta_0,X_i)]
$$

and the problem comes from the bound of $|nW_{in}-\gamma(T_i)|$ given by Lemma A.1 which contains $C(T_i)^{\alpha+\eta}$ (with $\eta>0$), a quantity that is not square integrable if we need to take $\alpha=1/2$. To show the negligibility of $\tilde{Q}_{n11}^{\beta}$ to $\tilde{Q}_{n13}^{\beta}$, apply Lemma A.1 with $\alpha=1/2$ and η equal to ρ from Assumption 7, and use Taylor expansion to bound $|f(\hat{\theta},X_j)-f(\theta_0,X_j)|$ by a constant times $\|\hat{\theta}-\theta_0\|$. Hence, $\tilde{Q}_{n11}^{\beta}$ to $\tilde{Q}_{n13}^{\beta}$ are bounded by

$$
O_P(n^{-1})\times\sum_{i\neq j}\frac{\gamma(T_i)\,|T_i-\beta f(\theta_0,X_i)|}{[C(T_i)]^{-(1/2+\rho)}}\gamma^{\beta}(T_j)\,a_{ij}(h)=O_P(n^{-1})\times B_{n1},
$$

$$
O_P(n^{-1})\times\sum_{i\neq j}\frac{\gamma(T_i)}{[C(T_i)]^{-(1/2+\rho)}}\gamma^{\beta}(T_j)\,a_{ij}(h)=O_P(n^{-1})\times B_{n2},
$$

and

$$
O_P(n^{-1})\times\sum_{i\neq j}\frac{\gamma(T_i)a_{ij}(h)}{[C(T_i)]^{-(1/2+\rho)}}\left(\frac{\hat{G}(T_j-)-G(T_j)}{1-G(T_j)}\gamma(T_j)\right)^{\beta}=O_P(n^{-1})\times B_{n3},
$$

respectively. It is easy to see that $\mathbb{E}(B_{nj})\leq c$, $j=1,2,3$, for some constant c independent of n and $h\in\mathcal{H}_n$. Deduce that for $j=1,2,3$, $B_{nj}=O_P(1)$. Lopez and Patilea (2006) showed that these orders hold uniformly in $h\in\mathcal{H}_n$. Collecting results, $\sup_{h\in\mathcal{H}_n}h^{\gamma}|Q_{n1}^{\beta}(\hat{\theta},\theta_0)|=O_P(n^{-1})$. ■

Lemma A.5 *Let the assumptions of Theorem 4.1 hold true. If* $\tau < \tau_H$ *and*

$$Q_{n1}^{\beta}(\tau) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} \left[\hat{U}_i^{\beta} - U_i^{\beta}\right] \mathbf{1}_{\{T_i \leq \tau\}} U_j^{\beta} K_h (X_i - X_j), \quad \beta = 0, 1,$$

then for any $\zeta \in (0, 1/2)$, $\sup_{h \in \mathcal{H}_n} h^{\zeta} \left| Q_{n1}^{\beta}(\tau) \right| = O_P(n^{-1})$.

Proof. If $w_i^{\beta} = \delta_i [T_i - \beta f(\theta_0, X_i)] [1 - G(T_i)]^{-2}$ we can write $Q_{n1}^{\beta}(\tau) = Q_{n11}^{\beta}(\tau) + Q_{n12}^{\beta}(\tau)$ with

$$Q_{n11}^{\beta}(\tau) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} \left[\hat{G}(T_i-) - G(T_i)\right] \mathbf{1}_{\{T_i \leq \tau\}} w_i^{\beta} U_j^{\beta} K_h (X_i - X_j)$$

$$Q_{n12}^{\beta}(\tau) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} \frac{\left[\hat{G}(T_i-) - G(T_i)\right]^2}{1 - \hat{G}(T_i-)} \mathbf{1}_{\{T_i \leq \tau\}} w_i^{\beta} U_j^{\beta} K_h (X_i - X_j).$$

By Theorem 2.1 of Gill (1983), $\sup_i [\hat{G}(T_i-) - G(T_i)]^2 \mathbf{1}_{\{T_i \leq \tau\}} = O_P(n^{-1})$. (The fact that the left endpoint of the support of the variables T_i may be $-\infty$ is of no consequence since we only consider $\hat{G}$ and G at the sample points.) Meanwhile, $\sup_{1 \leq i \leq n} G(T_i) \mathbf{1}_{\{T_i \leq \tau\}} \leq G(\tau) < 1$. These facts, Lemma A.2 and Assumption 4-(iv) imply

$$\sup_{h \in \mathcal{H}_n} \left| Q_{n12}^{\beta}(\tau) \right| \leq O_P(n^{-1}) \left(\frac{1}{n} \sum_{i=1}^{n} \left[w_i^{\beta}\right]^2 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^{n} \left[U_i^{\beta}\right]^2 \right)^{1/2} = O_P(n^{-1}).$$

To handle $Q_{n11}^{\beta}(\tau)$, we use the uniform i.i.d. representation of the Kaplan-Meier estimator, see Major and Rejtő (1988, Theorem 1):

$$\hat{G}(t-) - G(t) = \frac{1}{n} \sum_{k=1}^{n} \psi(T_k, t) + R_n(t)$$

with $\sup_{t \leq \tau} |R_n(t)| = O_P(n^{-1})$ and for each $t \leq \tau$,

$$\mathbb{E}[\psi(T_k, t)] = 0 \tag{A.5}$$

and $|\psi(T_k, t)| \leq M_1$ for some constant M_1 independent of t (but depending on τ). Now, we can write

$$\begin{aligned} Q_{n11}^{\beta}(\tau) &= \frac{1}{n^2(n-1)h^p} \sum_{i \neq j \neq k} \psi(T_k, T_i) \mathbf{1}_{\{T_i \leq \tau\}} w_i^{\beta} U_j^{\beta} K_h (X_i - X_j) \\ &+ \frac{1}{n} \frac{1}{n(n-1)h^p} \sum_{i \neq j} \psi(T_i, T_i) \mathbf{1}_{\{T_i \leq \tau\}} w_i^{\beta} U_j^{\beta} K_h (X_i - X_j) \\ &+ \frac{1}{n} \frac{1}{n(n-1)h^p} \sum_{i \neq j} \mathbf{1}_{\{T_i \leq \tau\}} w_i^{\beta} \psi(T_j, T_j) U_j^{\beta} K_h (X_i - X_j) + \{\text{remainder}\} \\ &= (n-2) n^{-1} Q_{n111}^{\beta}(\tau) + n^{-1} Q_{n112}^{\beta}(\tau) + n^{-1} Q_{n113}^{\beta}(\tau) + O_P(n^{-1}). \end{aligned}$$

By Lemma A.2, the fact that $\psi(\cdot,\cdot)$ is bounded and w_i^β, U_j^β are square integrable,

$$\sup_{h\in\mathcal{H}_n}\left\{\left|Q_{n112}^\beta(\tau)\right|+\left|Q_{n113}^\beta(\tau)\right|\right\}=O_P(1).$$

For $Q_{n111}^\beta(\tau)$, which is a $U-$process of order 3, apply the Hoeffding decomposition and write it as the sum of two degenerate $U-$processes

$$Q_{n1111}^\beta(\tau)=Q_{n111}^\beta(\tau)-Q_{n1112}^\beta(\tau)$$

and $Q_{n1112}^\beta(\tau)=n^{-1}(n-1)^{-1}\sum_{j\neq k}\phi_{jk}U_j^\beta$, where

$$\phi_{jk}=\mathbb{E}\left[\psi(T_k,T_i)\mathbf{1}_{\{T_i\leq\tau\}}w_i^\beta h^{-p}K_h(X_i-X_j)\mid X_j,T_k\right].$$

Notice that $|\phi_{jk}|\leq M_2$ for some constant M_2. The fact that $\mathbb{E}\left[U_j^\beta\mid X_j\right]=0$ a.s. and the property (A.5) make that the other terms in the Hoeffding decomposition of $Q_{n111}^\beta(\tau)$ are null. Corollary 4 of Sherman (1994) implies $\sup_{h\in\mathcal{H}_n}h^p\left|Q_{n1111}^\beta(\tau)\right|=O_P(n^{-3/2})$. Thus

$$\sup_{h\in\mathcal{H}_n}\left|Q_{n1111}^\beta(\tau)\right|=o_P(n^{-1}).$$

Next, fix $\zeta\in(0,1/2)$ and $\alpha\in(0,1)$ such that $1-\zeta/p<\alpha$, and consider the intervals H_l like in the proof of our Lemma A.4. For each H_l, by Sherman's (1994) Main Corollary

$$\begin{aligned}\mathbb{E}\left[\sup_{h\in H_l}|nh^\zeta Q_{n1112}^\beta(\tau)|\right] &\leq h_l^{\zeta-p}\mathbb{E}\left[\sup_{h\in H_l}|nh^pQ_{n1112}^\beta(\tau)|\right]\\ &\leq\Lambda_1h_l^{\zeta-p}\left[\mathbb{E}\sup_{h\in H_l}\left\{\frac{h^{2p}}{4n^2}\sum_{1\leq j,k\leq 2n}\phi_{jk}^2\left[U_j^\beta\right]^2\right\}^\alpha\right]^{1/2}\\ &\leq\Lambda_2h_l^{\zeta-(1-\alpha)p}\left(\frac{h_{l-1}}{h_l}\right)^{\alpha p}\left[\frac{1}{2n}\sum_{j=1}^{2n}\left[U_j^\beta\right]^2\right]^{\alpha/2}\\ &=h_{max}^{a_l}O_P(1),\end{aligned}$$

where Λ_1,Λ_2 are constants and a_l is like in the proof of Lemma A.4. Finally, sum over all l to obtain $nh^\zeta Q_{n1112}^\beta(\tau)=o_P(1)$ uniformly in $h\in\mathcal{H}_n$. ∎

Lemma A.6 *Let the assumptions of Theorem 4.1 hold true and let*

$$Q_{n2}^\beta=\frac{1}{n(n-1)h^p}\sum_{i\neq j}\left[\hat{U}_i^\beta-U_i^\beta\right]\left[\hat{U}_j^\beta-U_j^\beta\right]K_h(X_i-X_j),\quad\beta=0,1.$$

Then $\sup_{h\in\mathcal{H}_n}\left|Q_{n2}^\beta\right|=O_P(n^{-1})$.

Proof. Apply Lemma A.1 with $\alpha = 1/2$ to bound $|\hat{U}_i^\beta - U_i^\beta|$. Then,

$$\left|Q_{n2}^\beta\right| \leq \frac{O_P(n^{-1})}{n(n-1)} \sum_{i \neq j} \frac{\{|T_i|+1\}\gamma(T_i)}{[C(T_i)]^{-(1/2+\rho)}} h^{-p} K_h(X_i - X_j) \frac{\{|T_j|+1\}\gamma(T_j)}{[C(T_j)]^{-(1/2+\rho)}}.$$

By (2.3) and taking conditional expectations, the expectation of a term in the sum is

$$\mathbb{E}\left[\frac{\{|Y_1|+1\}}{[C(Y_1)]^{-(1/2+\rho)}} h^{-p} K_h(X_1 - X_2) \frac{\{|Y_2|+1\}}{[C(Y_2)]^{-(1/2+\rho)}}\right] = \mathbb{E}\left[q_\rho(X_1) q_\rho(X_2) h^{-p} K_h(X_1 - X_2)\right].$$

Since the last expectation is bounded, deduce that $Q_{n2}^\beta = O_P(n^{-1})$. Moreover, this rate holds uniformly in $h \in \mathcal{H}_n$, see Lopez and Patilea (2006) for the details. ∎

Lemma A.7 *Let Q_{n1}^β and $Q_{n1}^\beta(\tau)$ be defined as in (A.7) and (A.7), respectively. Under the assumptions of Theorem 4.1, for $\beta = 0$ or 1*

$$\sup_{h \in \mathcal{H}_n} h^{p/2} \left|Q_{n1}^\beta(\tau) - Q_{n1}^\beta\right| = C_\tau \times O_P(n^{-1}),$$

with the $O_P(n^{-1})$ factor independent of τ and $C_\tau \to 0$ when $\tau \uparrow \tau_H$.

Proof. Decompose

$$\frac{n-1}{n} h^{p/2} [Q_{n1}^\beta(\tau) - Q_{n1}^\beta] = \frac{1}{n^2 h^{p/2}} \sum_{1 \leq i,j \leq n} U_i^\beta K_h(X_i - X_j) \left(U_j^\beta - \hat{U}_j^\beta\right) \mathbf{1}_{\{T_j > \tau\}}$$
$$- \frac{K(0)}{n^2 h^{p/2}} \sum_{j=1}^n U_j^\beta \left(U_j^\beta - \hat{U}_j^\beta\right) \mathbf{1}_{\{T_j > \tau\}} = S_1 - S_2.$$

By the inverse Fourier transform and Cauchy-Schwarz inequality

$$|S_1| \leq \left[\int \hat{K}(hu) \left|\frac{1}{n} \sum_{j=1}^n \left(U_j^\beta - \hat{U}_j^\beta\right) \exp(2i\pi u' X_j) \mathbf{1}_{\{T_j > \tau\}}\right|^2 du\right]^{1/2}$$
$$\times \left[h^p \int \hat{K}(hu) \left|\frac{1}{n} \sum_{j=1}^n U_j^\beta \exp(-2i\pi u' X_j)\right|^2 du\right]^{1/2} = [S_{11}]^{1/2} [S_{12}]^{1/2}.$$

By the monotonicity of $\hat{\hat{K}}$, to obtain the uniform rate for S_{11} it suffices to take $h = h_{min}$ (see also Lemma A.2 in Lopez and Patilea, 2006). Now, by the Fourier transform,

$$S_{11} = \frac{1}{n^2 h_{min}^p} \sum_{i \neq j} (U_i^\beta - \hat{U}_i^\beta) \mathbf{1}_{\{T_i > \tau\}} K_{h_{min}}(X_i - X_j) (U_j^\beta - \hat{U}_j^\beta) \mathbf{1}_{\{T_j > \tau\}}$$
$$+ \frac{K(0)}{n^2 h_{min}^p} \sum_{j=1}^n \left(U_j^\beta - \hat{U}_j^\beta\right)^2 \mathbf{1}_{\{T_j > \tau\}} = S_{111} + S_{112}.$$

To handle S_{111}, apply Lemma A.1 with $\alpha = 1/2$. Then, $|S_{111}|$ is bounded by

$$\frac{O_P(n^{-1})}{n^2 h_{min}^p} \sum_{i\neq j} \frac{\{|T_i|+1\}\mathbf{1}_{\{T_i>\tau\}}\gamma(T_i)}{[C(T_i)]^{-(1/2+\rho)}} K_{h_{min}}(X_i - X_j) \frac{\{|T_j|+1\}\mathbf{1}_{\{T_j>\tau\}}\gamma(T_j)}{[C(T_j)]^{-(1/2+\rho)}},$$

where the $O_P(n^{-1})$ rate does not depend on τ. By (2.3) and taking conditional expectations, the expectation of a term in the last sum is

$$\mathbb{E}\left[\frac{\{|Y_1|+1\}\mathbf{1}_{\{Y_1>\tau\}}}{[C(Y_1)]^{-(1/2+\rho)}} K_{h_{min}}(X_1 - X_2) \frac{\{|Y_2|+1\}\mathbf{1}_{\{Y_2>\tau\}}}{[C(Y_2)]^{-(1/2+\rho)}}\right]$$
$$= \mathbb{E}\left[q_{\rho,\tau}(X_1) q_{\rho,\tau}(X_2) K_{h_{min}}(X_1 - X_2)\right] \to 0, \quad \text{when } \tau \uparrow \tau_H,$$

where $q_{\rho,\tau}(x) = \mathbb{E}[\{|Y|+1\}\mathbf{1}_{\{Y>\tau\}} C(Y)^{1/2+\rho} \mid X = x]$. Consequently, $|S_{111}|$ is bounded by $C_\tau \times O_P(n^{-1})$ for some C_τ independent of n but tending to zero as $\tau \uparrow \tau_H$. Next, to bound S_{112}, apply Lemma A.1 with $\alpha = 1/6$. Then

$$\begin{aligned} |S_{112}| &\leq \frac{1}{n^2 h_{min}^p} \sum_{j=1}^n \left(U_j^\beta - \hat{U}_j^\beta\right)^2 \mathbf{1}_{\{T_j>\tau\}} K(0) \qquad \text{(A.6)} \\ &\leq n^{-1/3} h_{min}^{-p} O_P(n^{-1}) \frac{1}{n} \sum_{j=1}^n \frac{\gamma(T_j)^2 \{|T_j|+1\}^2}{[C(T_j)]^{-(1/3+2\rho/3)}}. \end{aligned}$$

By Hölder inequality, the expectation of the last empirical mean is bounded by

$$\mathbb{E}^{1/3}\left[\delta\{|T|+1\}^4 [1-G(T)]^{-3}\right] \mathbb{E}^{2/3}\left[\{|T|+1\} C(T)^{1/2+\rho}\right],$$

which is finite under Assumptions 4-(iv) and 7. Finally, recall that $n h_{min}^{3p} \to \infty$. Collecting results, $\sup_{h\in\mathcal{H}_n} S_{11} = C_\tau \times O_P(n^{-1})$. To handle S_{12}, by the inverse Fourier transform and Corollary 4 of Sherman (1994) we obtain

$$S_{12} = \frac{1}{n^2} \sum_{i\neq j} U_i^\beta U_j^\beta K_h(X_i - X_j) + \frac{K(0)}{n^2} \sum_{j=1}^n \left[U_j^\beta\right]^2 = O_P(n^{-1}),$$

and the rate $O_P(n^{-1})$ is uniform in $h \in \mathcal{H}_n$. For S_2, take absolute values, apply Lemma A.1 with $\alpha = 1/4$ and use $n^{1/4} h_{min}^{p/2} \to \infty$ to deduce $\sup_{h\in\mathcal{H}_n} |h^{p/2} S_2| = o_P(n^{-1})$. ■

Proof of Theorem 4.1. *Step 1.* First, the assumptions ensure $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$ (see, e.g., Delecroix *et al.* 2006). Next, by Lemma A.4

$$\sup_{h\in\mathcal{H}_n} h^{p/2} \left|Q_n^\beta(\hat{\theta}) - Q_n^\beta(\theta_0)\right| = o_P(n^{-1}),$$

and thus we reduce the original problem to the study of $Q_n^{\beta}(\theta_0)$.

Step 2. Let us simplify notation: for $\beta = 0$ or 1 and $i = 1, ..., n$, write U_i^{β} (resp. $\hat{U}_i^{\beta}$) instead of $U_i^{\beta}(\theta_0)$ (resp. $\hat{U}_i^{\beta}(\theta_0)$). Now, decompose $Q_n^{\beta}(\theta_0) = \tilde{Q}_n^{\beta}(\theta_0) + 2Q_{n1}^{\beta} + Q_{n2}^{\beta}$ where

$$\begin{aligned}
Q_{n1}^{\beta} &= \frac{1}{n(n-1)h^p} \sum_{i \neq j} \left[\hat{U}_i^{\beta} - U_i^{\beta}\right] U_j^{\beta} K_h(X_i - X_j) \\
Q_{n2}^{\beta} &= \frac{1}{n(n-1)h^p} \sum_{i \neq j} \left[\hat{U}_i^{\beta} - U_i^{\beta}\right] \left[\hat{U}_j^{\beta} - U_j^{\beta}\right] K_h(X_i - X_j).
\end{aligned}$$

Fix $\tau < \tau_H = \inf\{t : H(t) = 1\}$ arbitrarily. To show that Q_{n1}^{β} is negligible, first we study a truncated version of this quantity, that is

$$Q_{n1}^{\beta}(\tau) = \frac{1}{n(n-1)h^p} \sum_{i \neq j} \left[\hat{U}_i^{\beta} - U_i^{\beta}\right] \mathbf{1}_{\{T_i \leq \tau\}} U_j^{\beta} K_h(X_i - X_j). \tag{A.7}$$

By Lemma A.5

$$\sup_{h \in \mathcal{H}_n} \left| h^{p/2} Q_{n1}^{\beta}(\tau) \right| = o_P(n^{-1}). \tag{A.8}$$

Step 3. Since $Q_{n1}^{\beta}(\tau_H) = Q_{n1}^{\beta}$, it remains to make $\tau \uparrow \tau_H$. By Lemma A.7,

$$\sup_{h \in \mathcal{H}_n} h^{p/2} \left| Q_{n1}^{\beta}(\tau) - Q_{n1}^{\beta} \right| = C_{\tau} \times O_P(n^{-1}),$$

with the $O_P(n^{-1})$ factor independent of τ and C_{τ} tending to zero when $\tau \uparrow \tau_H$. From the Cramér-Slutsky argument from Theorem 1.1 of Stute (1995), deduce that

$$\sup_{h \in \mathcal{H}_n} \left| n h^{p/2} Q_{n1}^{\beta} \right| = o_P(1),$$

which leads to $\sup_{h \in \mathcal{H}_n} \left| n h^{p/2} Q_n^{\beta}(\theta_0) - n h^{p/2} \tilde{Q}_n^{\beta}(\theta_0) \right| = o_P(1)$.

Step 4. Using arguments like those used in the previous proofs, it can be shown that under H_0, for $\beta = 0$ or 1,

$$\sup_{h \in \mathcal{H}_n} \left| \tilde{V}_n^{\beta}(\theta_0) / \hat{V}_n^{\beta} - 1 \right| = o_P(1).$$

See Lopez and Patilea for the details. This completes the proof of the first part of the theorem. The second part follows easily since $\tilde{V}_n^{\beta}(\theta_0)$ converges in probability to a strictly positive limit and $n h^{p/2} \tilde{Q}_n^{\beta}(\theta_0)$ is bounded in probability. ■

Proof of Lemma 4.3. For $i = 1, ..., n$, let

$$\begin{aligned} U_{in}^0 &= \frac{\delta_{in} T_{in}}{1 - G(T_{in})} - f(\theta, X_i), \qquad \hat{U}_{in}^0 = \frac{\delta_{in} T_{in}}{1 - \hat{G}(T_{in}-)} - f(\theta, X_i), \\ U_{in}^1 &= \frac{\delta_{in} \left[T_{in} - f(\theta, X_i)\right]}{1 - G(T_{in})}, \qquad \hat{U}_{in}^1 = \frac{\delta_{in} \left[T_{in} - f(\theta, X_i)\right]}{1 - \hat{G}(T_{in}-)}, \end{aligned}$$

By Lemma A.1 applied with $\alpha = 1/2$ and the boundedness of $f(\cdot, \cdot)$, for $\beta = 0$ or 1

$$|\hat{U}_{in}^\beta - U_{in}^\beta| = |R_{in}^\beta| \leq O_P(n^{-1/2}) \frac{\delta_{in}}{1 - G\left(T_{in}\right)} \{|T_{in}| + 1\} [C^{(n)}\left(T_{in}\right)]^{1/2+\eta}.$$

Now, simplify the notation $K_h(X_i - X_j)$ to K_{ij} and write

$$\begin{aligned} &\frac{1}{n^2 h^p} \sum_{i \neq j} \left\{ \hat{U}_{in}^\beta \hat{U}_{jn}^\beta - U_{in}^\beta U_{jn}^\beta \right\} K_{ij} = \frac{2}{n^2 h^p} \sum_{i \neq j} R_{in}^\beta U_{jn}^\beta K_{ij} + \frac{1}{n^2 h^p} \sum_{i \neq j} R_{in}^\beta R_{jn}^\beta K_{ij} \\ &= 2 \int \hat{K}\left(hu\right) \left(\frac{1}{n} \sum_{j=1}^n U_{jn}^\beta \exp\left(2i\pi u' X_j\right) \right) \left(\frac{1}{n} \sum_{j=1}^n R_{jn}^\beta \exp\left(-2i\pi u' X_j\right) \right) du \\ &\quad - \frac{2K\left(0\right)}{n^2 h^p} \sum_{j=1}^n R_{jn}^\beta U_{jn}^\beta \\ &\quad + \int \hat{K}\left(hu\right) \left| \frac{1}{n} \sum_{j=1}^n R_{jn}^\beta \exp\left(2i\pi u' X_j\right) \right|^2 du - \frac{K\left(0\right)}{n^2 h^p} \sum_{j=1}^n [R_{jn}^\beta]^2. \end{aligned}$$

The first integral can be bounded using Cauchy-Schwarz inequality and the bound of the second integral. To show that the second integral is of order $O_P(n^{-1})$, apply Lemma A.1 with $\alpha = 1/2$ and check that the expectation

$$\mathbb{E}\left[\frac{1}{h^p} \; K_{12} \; \frac{\gamma(T_{1n})\{|T_{1n}| + 1\}}{[C^{(n)}\left(T_{1n}\right)]^{-(1/2+\eta)}} \; \frac{\gamma(T_{2n})\{|T_{2n}| + 1\}}{[C^{(n)}\left(T_{2n}\right)]^{-(1/2+\eta)}} \right] \tag{A.9}$$

is bounded, where $\gamma(T_{1n}) = \delta_{1n}[1 - G\left(T_{1n}\right)]^{-1}$. From Assumption 8-(ii), deduce that this expectation equals

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{h^p} K_{12} \mathbb{E}\left[\frac{|Y_{1n}| + 1}{[C^{(n)}\left(Y_{1n}\right)]^{-(1/2+\eta)}} \mid X_1 \right] \mathbb{E}\left[\frac{|Y_{2n}| + 1}{[C^{(n)}\left(Y_{2n}\right)]^{-(1/2+\eta)}} \mid X_2 \right] \right] \\ &= \mathbb{E}\left[h^{-p} K_{12} q_\rho^{(n)}(X_1) q_\rho^{(n)}(X_2) \right] \end{aligned}$$

and the last expectation is bounded by Assumption 9. The rest of the proof continues with obvious arguments. ■

The proof of the following lemma is quite standard and is therefore omitted. It can be found in Lopez and Patilea (2006).

Lemma A.8 *Let Assumptions 4-(i) to (iii), 5, 6, 8, 9 hold true and let $\hat{\theta}$ denote either θ^{SD} or θ^{WLS}.*

i) If for all $n \geq 1$, $\mathbb{E}\left[\lambda_n(X)\nabla_\theta f\left(\theta_0, X\right)\right] = 0$ and $0 \leq |\lambda_n(\cdot)| \leq M_\lambda < \infty$ for some constant M_λ and if $\mathbb{E}\left|\lambda_n(X)\right| \to 0$, under the alternatives H_{1n} defined in (4.8), $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$.

ii) If Assumption 10 hold, under the alternative H_1, $\hat{\theta} - \bar{\theta} = O_P(n^{-1/2})$.

Proof of Theorem 4.4. Since $\sup_{\theta\in\Theta}\left|Q_n^\beta(\theta)\right|$ bounded in probability, Lemma 4.3 indicates that it remains to look at the limit of $\tilde{Q}_n^\beta(\hat{\theta})$. By Taylor expansion, arguments like those used in Lemma A.4 above and the fact that $\hat{\theta} - \bar{\theta} = O_P(n^{-1/2})$, we obtain $\sup_{h\in\mathcal{H}_n}\left|\tilde{Q}_n^\beta(\hat{\theta}) - \tilde{Q}_n^\beta(\bar{\theta})\right| = o_P(1)$. Now, since

$$U^\beta(\bar{\theta}) = \left\{\left[\gamma\left(T_i\right) - 1\right]\left[m(X_i) + \varepsilon_i - \beta f\left(\bar{\theta}, X_i\right)\right] + \varepsilon_i\right\} + \left\{m(X_i) - f\left(\bar{\theta}, X_i\right)\right\}$$

and $\mathbb{E}\left[\gamma\left(T_i\right) \mid X_i\right] = 1$, we can decompose $\tilde{Q}_n^\beta(\bar{\theta})$ in three parts: a degenerate and a zero-mean $U-$process of order 2 (indexed by h) that will negligible compared to the third part

$$\frac{1}{n(n-1)h^p}\sum_{i\neq j}\left[m(X_i) - f\left(\bar{\theta}, X_i\right)\right]\left[m(X_j) - f\left(\bar{\theta}, X_j\right)\right]K_h\left(X_i - X_j\right).$$

which tends to $\mathbb{E}\left[\left\{m(X) - f\left(\bar{\theta}, X\right)\right\}^2 g(X)\right] > 0$. Moreover, for $\beta = 0$ or 1 the variances $[\hat{V}_n^\beta]^2$ converge to

$$2\int K^2(u)du\,\mathbb{E}\left\{\mathbb{E}^2\left[U^\beta(\bar{\theta})^2 \mid X\right]g(X)\right\}, \tag{A.10}$$

uniformly in $h \in \mathcal{H}_n$. It is easy to see that for $\beta = 0$ or $\beta = 1$,

$$\begin{aligned}\mathbb{E}\left[U^\beta(\bar{\theta})^2 \mid X\right] &= \mathbb{E}\left[\left\{Y - \beta f(\bar{\theta}, X)\right\}^2\frac{G\left(Y\right)}{1 - G\left(Y\right)} \mid X\right] + \mathbb{E}\left[\varepsilon^2 \mid X\right] \\ &\quad + \left[m(X) - f\left(\bar{\theta}, X\right)\right]^2,\end{aligned}$$

and thus there is no general order between the limits in equation (A.10). ■

Proof of Theorems 4.5 and 4.6. Once again, Lemma 4.3 shows that we only need to look at $\tilde{Q}_n^\beta(\hat{\theta})$. Write $U_i^\beta(\theta) = u_{in} + v_{in} + w_{in} + \lambda_n\left(X_i\right) + \left\{f(\theta_0, X_i) - f\left(\theta, X_i\right)\right\}$ where

$u_{in} = [\gamma(T_{in}) - 1]\,\lambda_n(X_i)$,

$$\begin{aligned} v_{in} &= \beta\{\gamma(T_{in}) - 1\}\{f(\theta_0, X_i) - f(\theta, X_i)\}, \\ w_{in} &= \gamma(T_{in})\,\varepsilon_i + (1-\beta)\,[\gamma(T_{in}) - 1]\,f(\theta_0, X_i), \end{aligned}$$

and notice that $\mathbb{E}(u_{in} \mid X_i) = \mathbb{E}(v_{in} \mid X_i) = \mathbb{E}(w_{in} \mid X_i) = 0$ a.s. and there exists a sequence of real numbers $\overline{\sigma}_n^2$ tending to zero such that for each $n \geq 1$, $\mathbb{E}(u_{in}^2 \mid X_i) \leq \overline{\sigma}_n^2$. Using this decomposition of $U_i^\beta(\theta)$ we can split $\tilde{Q}_n^\beta(\hat{\theta})$ in several $U-$statistics of order 2. By repeated applications of Taylor expansion and Lemma A.3, and using the fact that $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$,

$$\begin{aligned} \tilde{Q}_n^\beta(\hat{\theta}) &= \frac{1}{n(n-1)} \sum_{i \neq j} w_{in} w_{jn} K_h(X_i - X_j) \\ &\quad + \frac{r_n^2}{n(n-1)} \sum_{i \neq j} \lambda(X_i)\,\lambda(X_j)\,\frac{1}{h^p} K_h(X_i - X_j) \\ &\quad + O_P\left(r_n n^{-1/2}\right) + o_P\left(n^{-1} h^{-p/2}\right), \end{aligned} \tag{A.11}$$

if $\lambda_n(\cdot) = r_n \lambda(\cdot)$. Moreover, since $\left| U_i^\beta(\hat{\theta}) - w_{in} \right| \leq o_P(1)\,[\gamma(T_{in}) + 1]$ with $o_P(1)$ independent of i,

$$\left[\hat{V}_n^\beta(\hat{\theta})\right]^2 - \frac{2}{n(n-1)\,h^p} \sum_{i \neq j}^{2} w_{in}^2 w_{jn}^2 K_h^2(X_i - X_j) = o_P(1). \tag{A.12}$$

From this and Lemma 2.1-(i) of Guerre and Lavergne (2005), the first $U-$statistic on the right-hand side of (A.11) multiplied by $nh^{p/2}$ and divided by $\hat{V}_n^\beta(\hat{\theta})$ converges in law to a standard normal distribution. Since the second $U-$statistic in (A.11) (without the r_n^2 factor) converges to $\mathbb{E}[\lambda^2(X) g(X)]$ in probability, and $\hat{V}_n^\beta(\hat{\theta})$ converges to a positive finite constant in probability, the proof of Theorem 4.5 is complete. Similarly, under the condition (4.10), $\tilde{Q}_n^{\beta}(\hat{\theta})$ can be decomposed

$$\begin{aligned}
\tilde{Q}_n^{\beta}(\hat{\theta}) &= \frac{1}{n\,(n-1)}\sum_{i\neq j} w_{in}w_{jn}K_h\,(X_i - X_j) \\
&\quad +(\hat{\theta}-\theta_0)'\frac{2}{n\,(n-1)}\sum_{i\neq j}\lambda_n\,(X_i)\,\nabla_\theta f(\theta_0, X_j)\frac{1}{h^p}K_h\,(X_i - X_j) \\
&\quad +\frac{2}{n\,(n-1)}\sum_{i\neq j}\lambda_n\,(X_i)\,w_{jn}\frac{1}{h^p}K_h\,(X_i - X_j) \\
&\quad +\frac{1}{n\,(n-1)}\sum_{i\neq j}\lambda_n\,(X_i)\lambda_n\,(X_j)\,\frac{1}{h^p}K_h\,(X_i - X_j) + \{\text{smaller terms}\} \\
&= \tilde{Q}_{na}^{\beta} + 2(\hat{\theta}-\theta_0)'\tilde{Q}_{nb}^{\beta} + 2\tilde{Q}_{nc}^{\beta} + \tilde{Q}_{nd}^{\beta} + \{\text{smaller terms}\}.
\end{aligned}$$

By Lemma A.3, $\tilde{Q}_{na}^{\beta} = O_P(n^{-1}h^{-p/2})$ and $|\tilde{Q}_{nc}^{\beta}| \leq O_P(n^{-1/2})\|\lambda_n\|_n$, while $|\tilde{Q}_{nb}^{\beta}| = O_P(1)\|\lambda_n\|_n$. Next, to obtain the rate of $\tilde{Q}_{nd}^{\beta}$, we follow the lines of the proof of Theorem 4 of Horowitz and Spokoiny (2001). See also Guerre and Lavergne (2005) and Lavergne and Patilea (2006). That is, approximating $\lambda_n(\cdot)$ by a piecewise polynomial function, we deduce

$$\tilde{Q}_{nd}^{\beta} \geq c\{1 + o_P(1)\}\left[\|\lambda_n\|_n - h^s\right]^2,$$

for some positive constant c, if $\lambda_n(\cdot) \in C(L, s)$ and the density $g(\cdot)$ is bounded away from zero. For the standard deviation, use (A.12) to deduce that $\hat{V}_n^{\beta}(\hat{\theta}) = O_P(1)$. Collecting results and taking h of order $n^{-2/(4s+p)}$, for any constant $c_1 > 0$, $\mathbb{P}(T_n^{\beta}(\hat{\theta}) > c_1) \to 1$ and this proves Theorem 4.6. ■

References

[1] Bose, A. and Sen, A. (2002). Asymptotic Distribution of the Kaplan–Meier U-Statistics. *J. Multivariate Anal.* **83**, 84–123.

[2] Delecroix, M., Lopez, O. and Patilea, V. (2006). Nonlinear censored regression using synthetic data. Working Paper No. 2006-10, CREST-INSEE.

[3] Gill, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49–58.

[4] Gray, R.J. and Pierce, D.A. (1985). Goodness-of-fit for censored data. *Ann. Statist.* **13**, 552–563.

[5] Guerre, E. and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33**, 840–870.

[6] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1296–1947.

[7] Hjort, N.L. (1990). Goodness of fit test in models for life history data based on cumulative hazard rates. *Ann. Statist.* **18** , 1221–1258.

[8] Horowitz, J.L. and Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599–631.

[9] Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9**, 1276–1288.

[10] Lavergne, P. and Patilea, V. (2006). Breaking the curse of dimensionality in nonparametric testing. Working Paper 2006-24, CREST-INSEE (www.crest.fr)

[11] Lawless, J.F. (2003). *Statistical Models and Methods for Lifetime Data,* 2nd ed. John Wiley & Sons, Inc., Hoboken.

[12] Lin, D.Y. and Spiekerman, C.F. (1996). Model checking techniques for parametric regression with censored data. *Scand. J. Statist.* **23**, 157–177.

[13] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301–309.

[14] Lopez, O. and Patilea, V. (2006). Nonparametric lack-of-fit tests for parametric mean-regression models with censored data. Working Paper 2007-01, CREST-INSEE.

[15] Major, P. and Rejtő, L. (1988). Strong Embedding of the Estimator of the Distribution Function under Random Censorship. *Ann. Statist.* **16**, 1113–1132.

[16] Miller, R. and Halpern, J. (1982). Regression with censored data. *Biometrika* **69**, 521–531.

[17] Nolan, D. and Pollard, D. (1987). $U-$processes : Rates of convergence. *Ann. Statist.* **15**, 780–799.

[18] Pardo-Fernández, J.C., van Keilegom, I. and González-Manteiga, W. (2005). Goodness-of-fit tests for parametric models in censored regression. DP 0519, Institut de Statistique, Louvain-la-Neuve (www.stat.ucl.ac.be)

[19] Rudin, W. (1987). *Real and complex analysis.* McGraw-Hill.

[20] Sánchez-Sellero, C., González-Manteiga, W. and van Keilegom, I. (2005). Uniform representation of product-limit integrals with applications. *Scand. J. Statist.* **32**, 563–581.

[21] Sherman, R.P. (1994). Maximal inequalities for degenerate $U-$processes with applications to optimization estimators. *Ann. Statist.* **22**, 439–459.

[22] Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal.* **45**, 89–103.

[23] Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422–439.

[24] Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.* **23**, 461–471.

[25] Stute, W. (1997). Nonparametric models checks for regression. *Ann. Statist.* **25**, 613–641.

[26] Stute, W. (1999). Nonlinear censored regression. *Statistica Sinica* **9**, 1089–1102.

[27] Stute, W., González-Manteiga, W. and Sánchez-Sellero, C. (2000). Nonparametric model checks in censored regression. *Comm. Statist. Theory Methods* **29**, 1611–1629.

[28] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New-York.

[29] Wei, L.J., Ying, Z. and Lin, D.Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.

[30] Zheng, Z. (1987). A class of estimators of the parameters in linear regression with censored data. *Acta Mathematicae Applicatae Sinica* **3**, 231–241.

[31] Zheng, J.X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75**, 263–289.

[32] Zhou, M. (1991). Some properties of the Kaplan-Meier estimator for independent nonidentically distributed random variables. *Ann. Statist.* **19**, 2266–2274.

[33] Zhou, M. (1992). M-estimation in censored linear models. *Biometrika* **79**, 837–841.